SEQUENCE VARIATION IN THE *APOA2* GENE AND ITS RELATIONSHIP WITH PLASMA HDL-CHOLESTEROL LEVELS

by

Sally Marie Hollister

B.S., University of Michigan, 2005

Submitted to the Graduate Faculty of

The Department of Human Genetics, Genetic Counseling

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

Sally Marie Hollister

It was defended on

April 9, 2008

and approved by

Thesis Advisor:
M. Ilyas Kamboh, Ph.D.
Professor and Acting Chair
Department of Human Genetics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
F. Yeşim Demirci, M.D.
Research Assistant Professor
Department of Human Genetics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
David N. Finegold, M.D.
Professor
Department of Pediatrics and Human Genetics
School of Medicine and Graduate School of Public Health
University of Pittsburgh

SEQUENCE VARIATION IN THE *APOA2* GENE AND ITS RELATIONSHIP WITH PLASMA HDL-CHOLESTEROL LEVELS

Sally Marie Hollister, M.S.

University of Pittsburgh, 2008

Coronary heart disease (CHD) is a major public health concern, affecting almost 16 million people in the U.S. and leading to 452,300 deaths in 2004 alone. Low levels of high density lipoprotein (HDL) cholesterol have been shown to increase the risk for cardiovascular disease (CVD). The role of genetics in affecting total cholesterol, HDL-cholesterol, and triglycerides levels is significant, with heritability estimates exceeding 50%. Recent studies have identified major loci associated with HDL-cholesterol through genome-wide association studies, which investigated influences of common variants on common traits. Relatively few studies have investigated the impact of rare variants on common disease. The aim of our study was to evaluate the role of genetic variation in *APOA2* (a biological candidate gene involved in HDL metabolism) in relation to HDL-cholesterol levels in epidemiological samples of African Blacks and U.S. Non-Hispanic Whites (NHWs). We resequenced the entire *APOA2* gene in individuals with HDL-cholesterol levels in the upper 5th percentile (47 NHWs and 48 Blacks) and lower 5th percentile (48 NHWs and 47 Blacks), allowing us to identify both rare and common variants. Common tagSNPs and all rare variants were screened in the larger NHW and Black samples for associations with HDL-cholesterol levels. We detected a total of 26 variants (25 single nucleotide substitutions and 1 microsatellite); 12 of which were previously unreported. Of the 12 new variants, 6 were present in NHWs and 6 in Blacks. We observed an increased number of minor alleles of *APOA2* variants (either increased heterozygosity for rare variants or increased homozygosity for common variants) in subjects with low HDL levels that was more pronounced

in NHWs. We performed a preliminary analysis using a total of 9 variants that were screened in NHWs (n=624, 8 variants) and Blacks (n=788, 5 variants) with TaqMan SNP genotyping assays to date. For the 8 variants that were screened in NHWs, we found significant association in only females for variants 2233C>T/rs6413453 (p=0.028) and 3251A>G (p=0.023). Completing genotyping for remaining variants will allow us to determine the extent to which *APOA2* variants influence HDL levels.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Dr. M. Ilyas Kamboh, for providing me with the opportunity to work on this project. I truly appreciate the time and energy you have invested for my sake. I would like to thank Dr. F. Yeşim Demirci, who has spent countless hours teaching, instructing, explaining, directing, and enlightening me, showing true enthusiasm for her work. Thank you for all the effort your have devoted, and for showing the utmost of patience with me! I would like to thank Dr. David N. Finegold, who has instilled the excitement of genetics in me and each student who has had the pleasure of having him as an instructor. I would also like to thank Dr. Candace M. Kammerer and Amy Dressen for their assistance in the statistical analysis for this project.

I would like to thank my coworkers and fellow students who have provided support, answered questions, and shared stories during my time in the Kamboh lab, especially Ryan Minster, Yuee Wang, Osama Khalifa, Sudeshna Dasgupta, Sangita Suresh, Jessica Figgins, and Sarah Hill.

I would like to thank the co-directors of the genetic counseling program, Ms. Elizabeth Gettig and Dr. Robin Grubs, who have worked tirelessly to provide a stimulating learning environment for the genetic counseling students.

I would like to thank my family and friends for your care and concern (especially my mom, who is graduating with her Master's in a few short months!). And thank you to Charlie, for all of your encouragement.

1.0 BACKGROUND AND SIGNIFICANCE

1.1 HEART DISEASE AND CHOLESTEROL LEVELS

Almost 16 million people in the U.S. are living with coronary heart disease (CHD), leading to 452,300 deaths in 2004 alone (Rosamond et al., 2007). The estimated direct and indirect cost of cardiovascular disease (CVD) in the U.S. for 2007 was $431.8 billion (Rosamond et al., 2007). Many factors influence the risk to develop heart disease including environmental factors like physical activity and diet and genetic factors like genes influencing lipid metabolism. Combining data from two twin studies, Berg et al. (1987) found that almost 50% of the variation of the lipid profile is under genetic control. Because of the significant contribution of genetics to CHD, understanding the underlying genetic influence of this disease is vital.

Risk factors for CHD include cigarette smoking, abnormal blood lipid levels, hypertension, diabetes, abdominal obesity, a lack of physical activity, low daily fruit and vegetable consumption, alcohol over-consumption, and psychosocial index (Yusuf et al., 2004). Hypertension is a particularly powerful risk factor for CHD in black persons, especially in black women (Rosamond et al., 2007).

Cholesterol levels are particularly important because both low levels of high-density lipoprotein (HDL) cholesterol and high levels of low-density lipoprotein (LDL) cholesterol have been shown to increase the risk for CVD. The American Heart Association website (2008)

classifies total cholesterol levels of 240mg/dL as high cholesterol. HDL-cholesterol levels of less than 40 mg/dL for men and less than 50 mg/dL for women is considered low HDL-cholesterol, which is a major risk factor for heart disease. An LDL-cholesterol level of 160-189 mg/dL is considered high, and each person's target level depends on his/her other risk factors, and could be as low as 100 mg/dL or less.

In 2004, 16.1% of Non-Hispanic White (NHW) males had total cholesterol levels of 240 mg/dL or higher. The percentage was 14.1% in Non-Hispanic Black (NHB) males, 18.2% in NHW females, and 12.5% in NHB (Rosamond et al., 2007). The prevalence of LDL-cholesterol levels of 130 mg/dL or higher in NHW males was 31.7%; 32.4% in NHB males; 33.8% in NHW females; and 29.8% in NHW females (Rosamond et al., 2007).

The difference between the sexes is most pronounced in the percentage of individuals with HDL-cholesterol levels less than 40 mg/dL. NHW males showed 26.2% in this category; 15.5% in NHB males. 8.8% in white females, and 6.9% in black females (Rosamond et al., 2007).

Clinically, it has been shown that lowering LDL (to under 70 mg/dL) is beneficial for individuals who have chronic CHD (Smith et al., 2006). HDL levels also influence CHD risk in men and women (Davidson et al., 2007). HDL levels are an independent risk factor for CHD. Modifying HDL-cholesterol levels "contributes significantly to both event rate reduction and changes in rates of atheromatous plaque progression (Nicholls, 2006; Brown, 2007).

Many studies have demonstrated the link between HDL levels and CVD. The Framingham Heart Study showed that there is an inverse relationship between CVD and HDL levels, meaning individuals with the highest HDL levels have the lowest risk for CVD (Castelli

et al., 1986). Each 1% increase in HDL-cholesterol was associated with a 2% reduction in the development of CHD (Castelli et al., 1986).

Low HDL levels have been shown to increase the risk of CHD in other studies. In the Physician's Health Study, participants with low HDL had an increased risk for CHD (Stampfer et al., 1991). Similar findings were demonstrated in the Israeli Ischemic Heart Disease Study, in which subjects with low total cholesterol and high HDL had the lowest rates of CHD-associated morbidity and mortality (Goldbourt et al., 1997). Individuals with less than 35 mg/dL HDL-cholesterol had a 3 times greater risk for CHD as compared to individuals with levels at 35 mg/dL or higher (Assman et al., 1998). Gordon et al. (1989) found that on average, each 1 mg/dL increase in HDL-cholesterol is associated with a 2% decrease in CHD risk in men and a 3% decrease in women.

A number of environmental factors are shown to have a negative affect on HDL-cholesterol levels (associated with decreased HDL-cholesterol levels), including Body Mass Index (BMI) and smoking. Other factors are correlated with increased HDL-cholesterol levels including moderate alcohol use, exercise, low sucrose and starch intake, and estrogen. (Women have higher HDL-cholesterol levels than men). Any of the many genes involved in HDL makeup, production and catabolism could have an affect on CHD as well.

1.2 GENETICS INFLUENCE ON HEART DISEASE

Estimates of the proportion of variation that can be attributed to inherited factors (heritability estimate) for total cholesterol, HDL-cholesterol, and triglycerides have exceeded

50% (Lusis et al., 2004). Researchers have attempted to identify these factors through twin studies, family studies and linkage analysis and, most recently, genome-wide association studies.

Major loci identified associated with HDL-cholesterol at genome-wide significance level in two recent studies were *ABCA1*, the *APOA1-APOC3-APOA4-APOA5* gene cluster, *CETP*, *LIPC*, *LIPG*, *LPL*, and *GALNT2* (Willer et al., 2008; Kathiresan el al., 2008). By applying multiple regression models during their meta-analysis; Willer et al. (2008) found that the variants identified in their study accounted for only ~5 to 8% of the variation in the lipid traits in their population. Comparing this number to the suggested heritability estimate for HDL-cholesterol, we can conclude that additional genetic factors remain to be identified, such as common variants with small effects and rare variants with large effects.

Many candidate genes have been proposed for research in relation to HDL levels in mice and in humans (Cambien 2005; Wang et al., 2006; Breslow 1988; Carlson et al. 2004, Humphries, 1988). These include biological and position candidate genes in a number of different categories, including HDL-associated apoliporoteins, HDL-associated enzymes and lipid transfer proteins, enzymes, cell receptors and transporters, and transcription factors such as. Since CHD is a complex disease, it is important to investigate each gene and associations of gene in their role(s) in contributing to HDL levels and heart disease.

1.3 HDL METABOLISM

The effect of HDL-cholesterol on CHD/CVD could occur by its levels (either by increasing synthesis or decreasing catabolism), or by increasing the availability of HDL-

cholesterol particles without necessarily increasing plasma HDL-cholesterol levels (Robinson et al. 2006; Toth et al., 2006).

The surface of HDL is made up of phospholipids, solubilized cholesterol, enzymes and apolipoproteins (Toth et al., 2006). Transcription factors, intracellular signaling pathways, cell surface receptors and enzymes all play a role in the regulation of serum HDL levels (Robinson et al., 2007; Toth et al., 2006; and Maxfield et al., 2002)

There are many different hypotheses attempting to explain the atheroprotective role of HDL-cholesterol. One hypothesis states that HDL-cholesterol exerts its influence on CHD risk through reverse cholesterol transport (Lewis et al., 2005). In reverse cholesterol transport, free cholesterol is removed from the peripheral macrophages (in the arterial wall) and is brought to the liver by HDL-cholesterol where it is excreted into bile (Toth, 2003). (Refer to Figure 1 for an illustration of reverse cholesterol transport.) HDL could reduce atherosclerosis in other ways. For example, methionine in apoA-I (carried in HDL's phospholipids surface), can reduce oxidized lipid species in LDL particles, making them anti-atherogenic (Toth, 2003)

HDL-cholesterol has been shown to perform a number of other functions leading to potential alterations of CHD risk including inhibiting expression of endothelial cell adhesion molecules, promoting endothelial nitric oxide synthesis and increasing vasodilation, and inhibiting pro-thrombotic and pro-inflammatory phenomena (Toth, 2007).

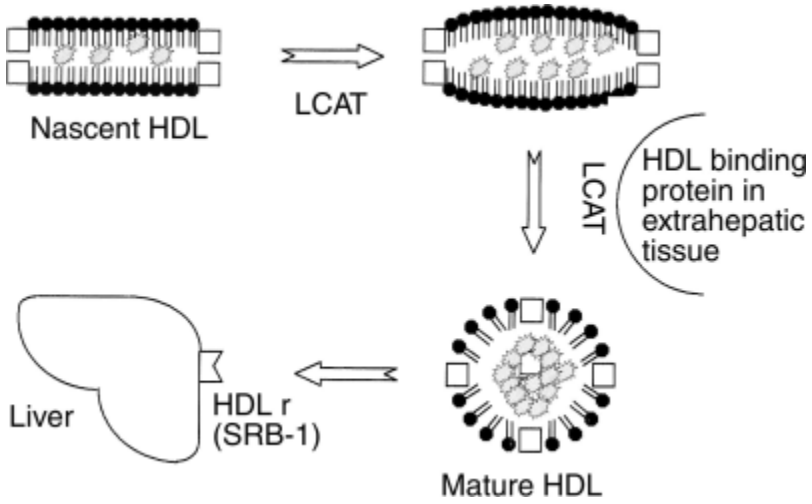


Figure 1. Reverse cholesterol transport from extra-hepatic tissues to the liver by high-density lipoprotein (HDL). HDLr: HDL receptor; LCAT, lecithin cholesterol acyltransferase (Vaziri, 2003).

1.4 HDL, APOA-II, AND LIPID METABOLISM

The apolipoprotein A-II (apoA-II) is the 2nd major protein constituent of HDL (Blanco-Vaca 2001). The mean apoA-II plasma concentration in normolipidemic humans is between 30-35 mg/dl (Bu et al., 1994). More than 20% of patients with CHD have apoA-II concentrations between 40 and 60 mg/dl (Vu-Dac et al., 1996; Bu et al., 1994). Mainly, apoA-II is found associated with HDL. A small fraction of apoA-II can be found associated with chylomicrons and VLDL (Alaupovic et al., 1991). ApoA-II may play an important role in the HDL remodeling process (Blanco-Vaca et al., 2001).

HDL is the smallest and most dense lipoprotein particle due to the high protein content (making up about 50% of HDL mass). ApoA-I and A-II are found in two main subpopulations of HDL; A-I-HDL (or Lp-I) contains only apoA-I, and A-I-A-II- HDL (or LpA-I-A-II) contains

both of the apolipoproteins). Both of these levels have been measured in recent epidemiologic studies (but have not been consistently measured in the past). Tailleux et al. (2002) conclude that apoA-I, LpA-I and apoA-II, LpAI-A-II are frequently low in patients with atherosclerosis, but that "the same relationship between A-II, LpAI-A-II, and atherosclerosis has not been perfectly established as with apoA-I and LpA-I. Tailleux et al. (2002) also conclude that apoA-II cannot compensate for apoA-I deficiency in patients with apoA-I deficiency (who show premature atherosclerosis even in the presence of remaining apoA-II.)

Barbaras et al. (1990) found that LpA-I may be more effective in capturing cholesterol from cells than Lp-A-I-A-II, suggesting that apoA-II may play an antagonistic role in cellular cholesterol efflux instead of sharing A-I's agonistic role (using proteins from purified mouse adipose cells.) Another study in reconstituted HDL showed that diacylglycerol was produced (a step in cholesterol efflux) when A-I but not A-II bound adipose cells (Theret et al., 1990), supporting the antagonistic role of apoA-II.

"The concentration of LpA1/A2 appears to be related to the synthesis of apoA-II in normolipidemic humans." (Ikewaki et al., 1995). This was supported by a mouse model with a polymorphism that affected the efficiency of translation of apoA-II. In these mice, the LpA2/LpA1 ratio increased (Doolittle et al., 1990).

Increasing the concentration of apoA-II in HDL showed a decrease in cholesterol efflux capacity in hepatoma cells in a study by Lagrost et al. (1995). Tailleux et al. (2002) report that this data confirms the role of apoA-II as an inhibitor of cellular efflux. Not all agree on an antagonistic role for apoA-II. A study using artificially reconstituted HDL with phospholipids and apolipoproteins found no difference in promotion of cholesterol efflux in HDL-cholesterol containing either apoA-I, A-II, or Cs (Mahlberg and Rothblat, 1992).

1.5 MOLECULAR ASPECTS OF APOA-II

ApoA-II is made in the liver as a 100 amino acid precursor (preproapo A-II) (Eggerman et al. 1991). This precursor is cleaved to proapo A-II (18 amino acids), and the mature apoA-II protein is produced with a final 5 amino acid cleavage (Brewer et al. 1986). The production of apoA-II is regulated by positive and negative regulatory elements from -911 to +29 relative to the cap site of the gene (Chambaz et al. 1991)

. The apoA-II species seen in plasma include the major homodimer formed by 2 polypetide chains of 77 amino acids linked by a disulfide bond at residue 6 (Brewer et al., 1972). Other minor species include the apoA-II-apoE heterodimer (in individuals with at least one *APOE*2* or *APOE*3* allele), the apoA-II-apoD heterodimer, and the apoA-II monomer (Brewer et al., 1972; Weisgraber and Mahley 1978; Borghini et al., 1991).

There are at least two lipid-associating domains in apoA-II (between amino acid 12 and 31 and 40 and 77 (Segrest et al., 1992). The amphipathic helices serve as lipid-associating protein detergent domains (Blanco-Vaca et al., 2001).

The NCBI reference protein sequence for apoA-II is #NM_001643 and is 100 amino acids in length. The NCBI RNA locus is #NM_001643, and is 473 nucleotides in length. *APOA2* has been mapped to chromosome 1 in region 1q21→1qter in humans (Li et al., 1988; Blanco-Vaca et al., 2001). ApoA-II is a member of the apolipoprotein superfamily which includes apoA-1, apoA-II, apoC's, and apoE. *APOA2* has 4 exons and 3 introns (see Figure 2) (Blanco-Vaca et al., 2001).

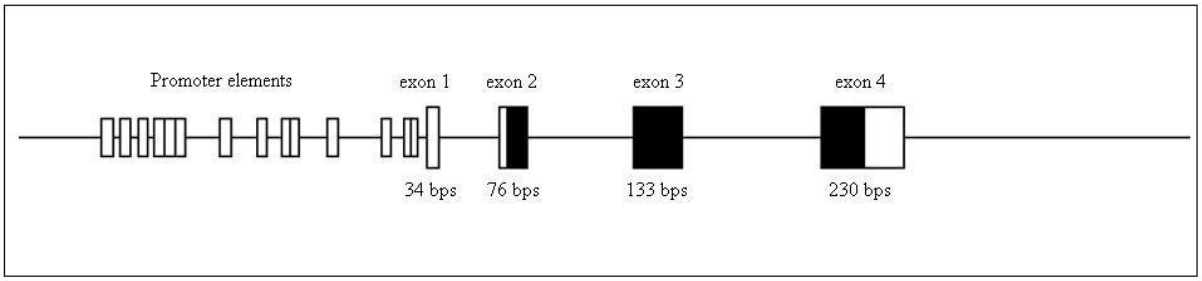


Figure 2. Structural organization of *APOA2*, including 4 exons and 3 introns. Figure adapted from Martin Campos et al. (2004).

1.6 *APOA2* POLYMORPHISMS

APOA2 is one of the 3 dozen genes involved in lipid metabolism (apolipoproteins, lipoprotein receptors, and lipid processing enzymes). The estimated heritability (h^2) for apoA-II is 0.69 (Berg, 1987). Studies of apoA-II in mice and humans have shown the antiatherogenic properties of apoA-II to be somewhat ambiguous.

Fourteen variants in the *APOA2* gene, plus the GT insertion/deletion polymorphism (close to the acceptor splice site of intron 2) have been reported by Fullerton et al. (2002) and summarized in the Seattle SNP Database for 3 populations (Jackson, MS; North Karelia, Finland; and Rochester, MN). A total of 4 variants (plus the GT repeat region) were reported in the Jackson population. (155C>G-MAF=0.04, 872C>T-MAF=0.33, 2233T>C-MAF=0.02, and 2868A>C-MAF=0.02). Thirteen variants (plus the GT polymorphism) were observed in the Finnish population (155C>G-MAF=0.25, 872C>T-MAF=-0.31, 1218C>G-MAF=0.02, 2038C>G-MAF=0.08, 2085T>C-MAF=0.02, 2115A>G-MAF=0.02, 2233T>C-MAF=0.15, 2818T>C-MAF=0.08, 2868A>C-MAF=0.15, 2994T>C-MAF=0.02, 3027C>T-MAF=0.10, 3092G>A-MAF=0.10, 3208A>G=MAF0.06). Twelve variants (plus the GT repeat region) were

observed in the Rochester population (155C>G-MAF=0.33, 201A>G-MAF=0.02, 872C>T-MAF=-0.38, 2038C>G-MAF=0.17, 2115A>G-MAF=0.02, 2233T>C-MAF=0.17, 2818T>C-MAF=0.17, 2868A>C-MAF=0.17, 2994T>C-MAF=0.15, 3027C>T-MAF=0.28, 3092G>A-MAF=0.29, 3208A>G=MAF0.12).

The GT repeat was identified by Weber and May, 1989. This polymorphism does not seem to have an effect on splicing (Blanco-Vaca et al., 2001). One SNP is a G-to-A change (termed *APOA2*$_{\text{Hiroshima}}$) (location 2115; rs6413452 in the current study) that affects the donor splice site of intron 3. Homozygosity for this allele is associated with familial apoA-II deficiency (Deeb et al., 1990). This change does not seem to affect heart disease risk. Another SNP observed by Dupuy-Gorce et al. (1996) is a C-to-T change that modifies a *Bst*NI restriction site near the acceptor splice site of intron 3 (2298C>T leu/leu synonymous change using Seattle Database nomenclature).

1.7 ASSOCIATION OF *APOA2* VARIANTS WITH PHENOTYPES

It is well established that apoA-I has a protective role in CVD. The role of genetic variation in *APOA2*, however, is not clearly understood. There is one case of human apoA-II deficiency, caused by a gene mutation at the splice junction of exon 3 (site 2115 G>A). This mutation is expected to block the splicing of intron 3 from the primary transcript, preventing functional mRNA from forming in homozygous individuals. In this family, the apoA-II deficiency does not seem to have significant impact on lipid levels or on CHD risk (Deeb et al., 1990).

Male patients with CHD have been studied for association with CA repeat region (position 1671), but it showed no association (Shadrina et al., 1997).

1.8 GENETIC VARIANTS AND COMMON TRAITS

Low HDL-cholesterol is a common trait in the general population and could be explained by common alleles in candidate loci, illustrating the "common trait-common variant" hypothesis. (Knoblauch et al., 2004; Morabia et al., 2003). Hegele et al. (1997) suggested that small genetic effects of common variants may be a more realistic genetic model for plasma lipoprotein variation. Other studies suggest that multiple rare variants are associated with low HDL-cholesterol levels, illustrating the "common trait-rare variant" hypothesis (Cohen et al., 2004).

Cohen et al. (2004) studied the effects of rare alleles in *ABCA1*, *APOA1* and *LCAT* genes on plasma HDL-cholesterol levels in the general population and found that nonsynonymous sequence variants were significantly more common in the low HDL-cholesterol group than in the high HDL-cholesterol group. (One SNP was found in the low HDL-cholesterol group that was not found in the high HDL-cholesterol group.)

Sequencing-based studies have been used to identify both the phenotypic effect of rare variants, and common variants with a modest effect. Crawford et al. (2006) have shown that by resequencing subjects with extreme quantitative traits, potentially medically relevant variants can be detected. They conclude that to fully understand the variety of alleles that affect human phenotypes, however, many different sequencing and genotyping methods may be necessary.

1.9 SPECIFIC AIMS

The objective of this study was to evaluate the genetic variation in the *APOA2* gene in relation to HDL-cholesterol levels in two well-characterized epidemiological samples of 788 African blacks from Benin City, Nigeria and 624 U.S. Non-Hispanic Whites (NHWs) from Colorado. Because *APOA2* is a biological candidate for HDL-cholesterol levels, its common and/or rare variants are hypothesized to contribute to variation in HDL-cholesterol levels.

Step 1: *APOA2* was resequenced in individuals with HDL-cholesterol in the upper 5th percentile (47 NHWs and 48 Blacks) and in the lower 5th percentile (48 NHWs and 47 Blacks). This allows us to identify rare variants in addition to common variants. The variants that are unique to one group or the other, or common to both groups that may be over-represented or under-represented in one group may have biological significance.

Step 2: Common tagSNPs and rare variants (MAF <5%) were screened in the entire NHW and Black population (individuals with varying levels of HDL-cholesterol). The effects of both rare and common variants on HDL-cholesterol levels were determined.

2.0 SUBJECTS AND METHODS

2.1 SUBJECTS

2.1.1 Sample Populations

Table 1 summarizes the subjects used in this study. Samples from the African Black population were recruited from Benin City, Nigeria as part of a study on CHD risk factors in Blacks. Demographic and health information was gathered from participants during the initial study. The subjects from Benin City were recruited from Junior and Senior staff in government ministries (representing different salary grades.) Detailed information about the study population can be found in Bunker et al. (1995, 1996).

Table 1. Sample populations

Population	**Men (%)**	**Women (%)**	**Total (%)**
African blacks	502 (62.4)	302 (37.6)	804 (53.8)
U.S. whites	322 (46.6)	369 (53.4)	691 (46.2)
Total	824 (55.1)	671 (44.9)	1,495 (100)

Samples from Non-Hispanic Whites (NHWs) were drawn from the San Luis Valley Southern Colorado Diabetes Study. All subjects who were included in the current study were normoglycemic and a more detailed description of the sample population can be found in Hamman et al. (1989) and Rewers et al. (1993).

Total HDL-cholesterol was determined enzymatically after dextran sulfate magnesium precipitation (Harris et. al 1998).

2.2 DNA SEQUENCING

DNA was extracted from clot samples (Black population) and from buffy coat (NHW population) using standard procedures. Individuals with serum HDL-cholesterol in the upper 5th percentile (47 white and 48 black) and in the lower 5th percentile (48 white and 47 black) were selected to be used to screen for common and rare variants by resequencing of DNA samples for the entire *APOA2*. Of 95 NHW individuals, 47 were females (23 with high HDL levels and 24 with low HDL levels); 48 were males (24 with high HDL levels and 24 with low HDL levels). Of 95 Black individuals, 48 were females (24 with high HDL levels and 24 with low HDL levels); 47 were males (24 with high HDL levels and 23 with low HDL levels). Refer to Table 2 for a summary of the population characteristics including age, BMI, LDL-, HDL-, and total cholesterol and triglyceride levels in NHWs and Blacks for both high HDL and low HDL subgroups.

Table 2. Biometric and quantitative data (mean±SD) of NHW and Black samples used for DNA sequencing.

Variable	*NHWs (n=95)*			*Blacks (n=95)*		
	High HDL (*n* =47)	Low HDL (*n* =48)	*p* -value	High HDL (*n* =48)	Low HDL (*n* =47)	*p* -value
Sex (M/F)	24/23	24/24	0.92	24/24	23/24	0.92
Age (years)	55.45 ± 9.80	53.03 ± 10.54	0.25	41.29 ± 8.72	40.87 ± 7.16	0.8
BMI (kg/m2)	23.17 ± 3.17	27.35±3.90	<0.001	22.06 ± 4.71	23.91 ± 5.51	0.08
Total cholesterol (mg/dl)	227.34±51.76	208.81±44.65	0.06	201 ± 39.68	141.68 ± 31.03	<0.001
LDL cholesterol (mg/dl)	126.84±46.95	136.95±41.28	0.28	112.55 ± 39.75	95.04 ± 28.28	0.02
HDL cholesterol (mg/dl)	**77.68±13.32**	**31.81±4.37**	**<0.001**	**76.05 ± 7.53**	**25.51 ± 5.66**	**<0.001**
Triglycerides (mg/dl)	114.09±60.88	240.21±153.22	<0.001	61.98 ± 19.85	95.79 ± 73.21	0.003

We used the publicly available information from the Seattle SNPs database (http://pga.mbt.washington.edu/) to order M13-tagged primers used to resequence *APOA2*. The M13-tagged primers that were used to produce overlapping resequencing amplicons are displayed in Table 3.

Table 3 *APOA2* Polymerase Chain Reaction (PCR) Primers

PCR Amplicons	Forward Primer	Reverse Primer
1	5'-tgtaaaacgacggccagtCAGCCTGAGCAACAAGAGCA-3'	5'-caggaaacagctatgaccCAGGAAGCAGGATTCCAAGTT-3'
2	5'-tgtaaaacgacggccagtTCAACCCTGTCCCTGATTTC-3'	5'-caggaaacagctatgaccCACCAGCACATTCTCCCTTT-3'
3	5'-tgtaaaacgacggccagtGTTGTGGAGAGGGAGAGAGC-3'	5'-caggaaacagctatgaccAGGCCCAGTTCTTGCTGTTC-3'
4	5'-tgtaaaacgacggccagtGCAAGGACCTGATGGAGAAG-3'	5'-caggaaacagctatgaccGCACCCTCATCCTGTCATTT-3'
5	5'-tgtaaaacgacggccagtCTAGAGCTCCTGTCCCTAC-3'	5'-caggaaacagctatgaccTGCCTGGTCTAGATAAGAGG-3'
Internal SEQ primer	5'-AAATGACAGGATGAGGGTGC-3'	

These primers produce 5 overlapping amplicons ranging in size from 811 to 906 base pairs (bps) to cover the entire gene, plus ~ 1.1 kb of 5' flanking region (putative promoter region) and ~ 900 bps of 3' flanking region. About 3.3 kb of genomic fragment were amplified and sequenced overall. Initially, all 5 amplicons were sequenced in one direction (forward). Subsequently, two amplicons were sequenced in the other direction (reverse) in order to fill the gaps. Despite both directional sequencing, a small gap was still observed for amplicon #5. In order to fill this small gap, we designed (Primer3 Software, version 0.4.0 http://Frodo.wi.mit.edu) and used an internal sequencing primer.

The PCR was performed on a thermal cycler with a heated lid (GeneAmp PCR System 9700, Applied Biosystems). The PCR (Polymerase Chain Reaction) reaction and cycling conditions are given in Table 4.

Table 4. PCR Reaction and Cycling Conditions

PCR Reaction (total volume 25 µL)		PCR conditions
DNA	3.0 µL	1. 95˚C for 5 minutes 2. 95˚C for 45 sec. 3. 61˚C for 45 sec. } 4. 72˚C for 1 min. (2-4 for 40 cycles) 5. 72˚C for 10 min. 6. Cool to 4˚C
dH_2O	11.75 µL	
10x BufferGold	2.5 µL	
$MgCl_2$ (25 mM)	3.0 µL	
dNTPs (1.25 mM)	3.8 µL	
Forward Primer (20mM)	0.4 µL	
Reverse Primer (20mM)	0.4 µL	
TaqPolymerase (5U/µL)	0.15 µL	

Gel electrophoresis was used to check for the success of the amplification of all PCR samples. For each sample, 7µL of PCR product was combined with 7µL of loading buffer and

6μL distilled water, and loaded into a 96-well pre-cast agarose gel (Invitrogen™ E-Gel® 96 2% with SYBR® Safe) and run for 8 minutes using the EG program on the electrophoresis base (Invitrogen™ E-Base™). For a small number of samples that failed initial amplification, reamplification and confirmation was performed using regular agarose gels. For this purpose, 7μL of PCR product was combined with 5μL of loading buffer and loaded into 2% agarose gel in TBE buffer (tris, boric acid, and disodium EDTA dihydrate) and stained with ethidium bromide (Amresco). Electrophoresis was run for about 25 minutes at 250V. Both 96-well pre-cast gels and regular agarose gels were visualized using a UV transilluminator.

The amplified samples were sent for capillary sequencing (Genomic Services of Agencourt Bioscience Corporation, Beverly, MD). Two programs were used to align and analyze the sequencing results; Sequencher version 4.7 (Gene Codes Corporation, Ann Arbor, MI), and Variant Reporter version 1.0 (Applied Biosystems, Foster City, CA).

2.3 GENOTYPING

2.3.1 Common SNPs

For common SNPs (MAF≥5%), we either ordered available pre-made TaqMan SNP genotyping assays or requested custom designed assays. Seven assays were available as pre-made TaqMan SNP genotyping assays. Two of four requested custom assays passed the quality control requirements and were used in our study. Table 5 details a total of 9 assays that were analyzed in either the African black or NHW population or both.

2.3.1.1 TaqMan analysis

Table 5. TaqMan® SNP Genotyping Assays

APOA2 reference SNP ID*	TaqMan® Assay Type	Assay ID	Population
rs3829793	Pre-made	C__27495574_10	Both
rs5082	Pre-made	C__11453334_10	Both
rs5087	Pre-made	C__11453973_20	NHWs
rs6413452	Pre-made	C__29505396_10	NHWs
rs6413453	Pre-made	C__25770132_10	Both
rs5088	Pre-made	C__11453980_10	Blacks
rs12721035	Pre-made	C__31430618_10	NHWs
rs12143180	Custom made	AoA2-2994	Both
3251	Custom made	ApA2-3251	NHWs

*Reference SNP ID number was used whenever the SNP was present in public databases. For newly identified variants, we used the same nomenclature as the SeattleSNPs database, based on the location in the reference DNA sequence (Accession # AY100524)

The TaqMan procedure involves amplification of the product and endpoint reading using the ABI Prism 7900HT Sequence Detection System (Applied Biosystems). TaqMan Genotyping Master Mix and Assay is added to 384-well plates containing dried whole genome amplified DNA. The Genotyping Assay contains sequence-specific forward and reverse primers, one TaqMan minor groove binder (MGB) probe labeled at the 5' end with VIC dye and one TaqMan MGB probe labeled at the 5' end with FAM dye to detect the alternative alleles. On the 3' end of the probes, a nonfluorescent quencher (NFQ) is attached. PCR amplification was done using a PTC-200 MJThermal Cycler (Biorad) or a GeneAmp 9700 (Applied Biosystems). The cycling conditions are displayed in Table 6.

Table 6. Thermal Cycler Conditions for TaqMan

Times and Temperatures		
Initial Step	Denaturation	Primer Annealing & Extension
HOLD	50 Cycles	
10 min @95°C	15 sec @95°C	1 min @ 60°C

Each TaqMan MGB probe anneals to the target sequence harboring the SNP of interest during the annealing step. Then, AmpliTaq Gold DNA polymerase cleaves the probes that hybridize to the target sequence. During this process, the reporter dye is separated from the quencher dye, releasing fluorescence. Fluorescence is suppressed if probes do not hybridize to the target sequence and the reporter dye does not separate from the quencher dye. Because of the selective annealing of the TaqMan MGB probes, discrimination of alleles is possible.

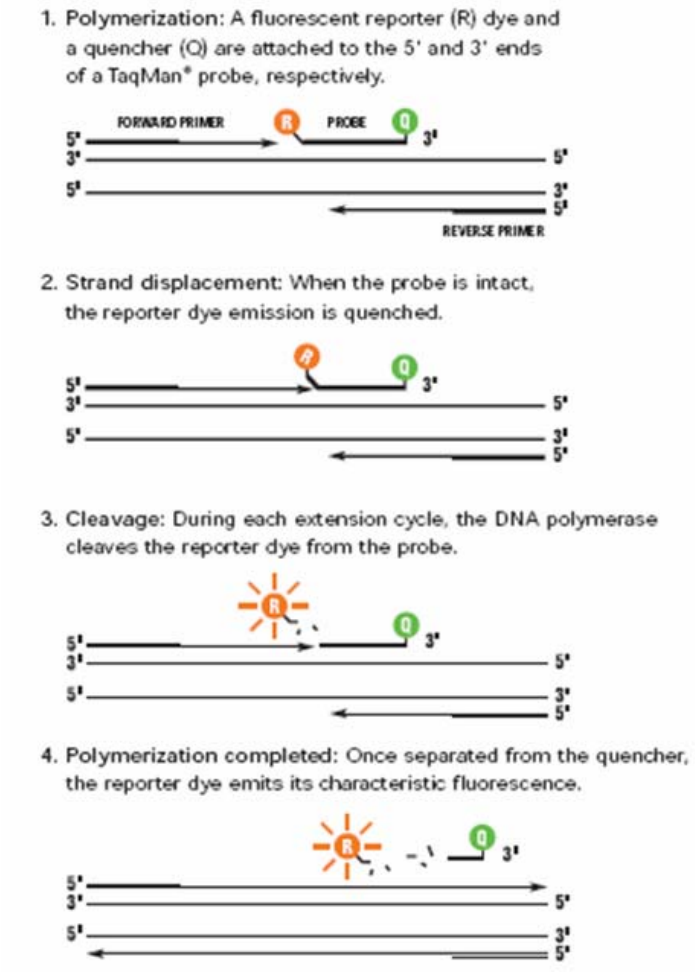


Figure 3. Illustration of TaqMan probe-target matches and mismatches. Wang et al., 2006.

2.3.2 Rare Variants

2.3.2.1 Restriction Fragment Length Polymorphism (RFLP)

Restriction Fragment Length Polymorphism (RFLP) is a technique than can be used to detect sequence alterations using restriction enzymes. Restriction enzymes or restriction endonucleases can recognize and cut DNA in the presence of a specific short sequence (usually a palindromic segment of 4-8 base pairs). The DNA can then be separated in gel electrophoresis,

allowing visualization of the fragments. In this way, the presence or absence of a specific nucleotide can be confirmed.

In our experiment, we searched for a possible RFLP site for all of our variants using NEBcutter V2.0 (New England Biolabs, Ipswitch, MA). We ordered RFLP primers and enzymes when the following conditions were met; a polymorphic restriction site concerning our target variant was present; there are not too many nearby non-polymorphic restriction sites that may hamper the gel-based allelic discrimination; the restriction site flanking sequences allow us to design successful PCR primers. Those three conditions were met for 4 variants (159, 336, 1055, and 1569).

The PCR primers that were used are displayed in Table 7. Reaction mixture quantities and conditions are given in Table 8. Agarose gel electrophoresis was used to confirm the successful amplification of PCR products using the same approach as for sequencing amplicons.

Table 7. RFLP PCR Primers

APOA2 variant	Forward Primer	Reverse Primer
159	5’-TGAGCAACAAGAGCAAAACTC-3’	5’-CAGAAGCTTTTGTATCTTTCACC-3’
336	5’-TTCACCCAAATGCTTTGAAC-3’	5’-TGGCAGGAAAAGAGGTGAAT-3’
1055	5’-CCATCACCATGAGTCTTCCA-3’	5’-TCTGTCCTTGGTGTCTGTGC-3’
1569	5’-AAAAGGGAGAATGTGCTGGT-3’	5’-GGTCCACAGCAGTGAATCCT-3’

Table 8. PCR Reaction and Cycling Conditions for RFLP

Reaction Conditions		Cycling Conditions (GeneAmp 9700 thermal cycler)
DNA	3.0 µL	95˚C for 5 minutes 95˚C for 45 sec. 60/61˚C for 45 sec. } 40 cycles 72˚C for 1 min. 72˚C for 10 min. Cool to 4˚C
dH_2O	12.25/11.75 µL	
10x BufferGold	2.5 µL	
$MgCl_2$ (25 mM)	2.5/3.0 µL	
dNTPs (1.25 mM)	3.8 µL	
Forward Primer	0.4 µL	
Reverse Primer	0.4 µL	
TaqPolymerase (5U/µL)	0.15 µL	

To confirm the variants that were detected by sequencing, we individually amplified the samples with known genotypes (all known heterozygotes and selected known wild-type homozygotes) and tested our RFLP method. After the confirmation of the variant and the success of the RFLP, we tested the pre- and post-PCR pooling strategy by combining different numbers of samples (either pre-PCR DNA or post-PCR amplicons) ranging from 5 to 10 samples. We used a higher volume of PCR for the digestion reaction when we pooled post-PCR products (between 20 and 24 µL of pooled sample was used).

From individual samples and pre-PCR pooled samples, we combined 17 µL of each sample with 13 µL of master mix including: 9.5µL dH_2O, 3 µL 10x buffer, and 0.5 µL of the restriction enzyme. The restriction enzyme and overnight incubation temperature used for 4 rare variants are given in Table 9.

Table 9. Restriction Enzyme and Overnight Temperature used in RFLP

***APOA2* variant**	**Enzyme used for digestion**	**Expected fragments**	**Overnight incubation temperature**
159C>G	*Cvi*KI-1	129+85 vs. 129+65+20	37°C
336T>G	*Msp*I	200 vs. 113+87	37°C
1055C>T	*Bsr*I	99+87 vs. 186	65°C
1569T>C	*Msp*I	218 vs.128+90	37°C

Fifteen µL of digested product was combined with 7µL loading buffer and loaded to 2% or 3% agarose gel (depending on the size difference of expected fragments) with ethidium bromide and electrophoresis was run for ~30-50 minutes at 150V.

2.4 STATISTICAL METHODS

Allele frequencies were determined by direct allele counting. Concordance of the genotype distribution to Hardy-Weinberg equilibrium was tested using a $\chi 2$ goodness-of-fit test for each variant. Fisher's exact test and standard Z-test of two binomial proportions were used to compare the genotype and allele frequencies, respectively. Linkage disequilibrium (LD) pattern and tagSNPs were determined using the Haploview version 3.32 (http://www.broad.mit.edu/mpg/haploview/). All dependent quantitative variables were transformed (using either log or square root transformation) to reduce the effects of non-normality whenever necessary. The significant covariates for each dependent variable were

identified using stepwise regression in both directions. The most parsimonious set of covariates was determined separately for males and females within each ethnic group. One-way analysis of variance (ANOVA) was performed separately for males and females within each ethnic group to test for the effects of genotypes on the means of the quantitative traits (transformed and adjusted when necessary). All computations were performed using the R statistical software package (version 2.3.1, http://www.r-project.org). A *p*-value of less than 0.05 was considered as suggestive evidence of association.

3.0 RESULTS

3.1 DNA SEQUENCING OF THE *APOA2* GENE

Complete *APOA2* sequencing was performed on DNA samples from 95 African Black and 95 American NHW individuals with the highest and lowest HDL levels. A total of 26 sequence variants were identified, including a microsatellite (dinucleotide repeat marker) in intron 2. Of the 26 variants, 14 were already reported in publicly available databases. Of the 25 single nucleotide substitutions (no indels detected), 9 were observed only in the NHWs, and 7 were observed only in Blacks (Table 10). Nine variants were seen in both NHWs and Blacks. The GT dinucleotide repeat marker was not evaluated in detail due to the difficulties in doing so using the software packages we had. Figure 4 is a schematic representation of the location of the variants identified in this study.

Among the 12 newly identified variants, only one had a MAF of 5% and the remaining 11 variants had less than 5% MAF. Seven variants had a MAF of less than 1%. Nine variants were located in the putative promoter region. Only one variant was in the coding region (exon 4, rs5088, 2298C>T), and was a silent mutation causing a synonymous amino acid change at position 83 (leucine/leucine). The remaining variants were in either intronic (8 variants including the microsatellite) or in the 3'-flanking regions (8 variants). Chromatograms illustrating the 12 new variants are shown in Figure 5 and Figure 6.

Table 10. *APOA2* variants identified in our study. *Location based on DNA reference sequence accession # AY100524 (the microsatellite marker is not included in this table.)

APOA2 Variant*	rs (CHIP&GB)	Population	Location	MAF(NHWs)	MAF(Blacks)
155G>C	rs3829793	Both	promoter	0.321	0.063
159C>G	new variant	NHWs	promoter	0.011	---
336T>G	new variant	Blacks	promoter	0.005	---
470G>A	new variant	Blacks	promoter	---	0.005
589C>T	new variant	Blacks	promoter	---	0.021
772C>T	new variant	NHWs	promoter	0.016	---
872T>C	rs5082	Both	promoter	0.405	0.195
898A>G	new variant	NHWs	promoter	0.005	---
1055C>T	new variant	Blacks	promoter	---	0.042
1569T>C	new variant	Blacks	intron 2	---	0.005
1638C>T	new variant	Blacks	intron 2	---	0.027
2038G>C	rs5085	Both	intron 3	0.168	0.095
2085C>T	rs5087	NHWs	intron 3	0.005	---
2115G>A	rs6413452	NHWs	intron 3	0.016	---
2218A>C	new variant	NHWs	intron 3	0.005	---
2233C>T	rs6413453	Both	intron 3	0.084	0.005
2298C>T	rs5088	Blacks	exon 4	---	0.026
2547G>T	new variant	Blacks	3' flanking	---	0.005
2818C>T	rs12143241	Both	3' flanking	0.168	0.095
2868C>A	rs12721036	Both	3' flanking	0.084	0.005
2994C>T	rs12143180	Both	3' flanking	0.163	0.053
3027T>C	rs685	Both	3' flanking	0.200	0.095
3092A>G	rs17244530	Both	3' flanking	0.197	0.083
3208G>A	rs12721035	NHWs	3' flanking	0.038	---
3251A>G	new variant	NHWs	3' flanking	0.048	---

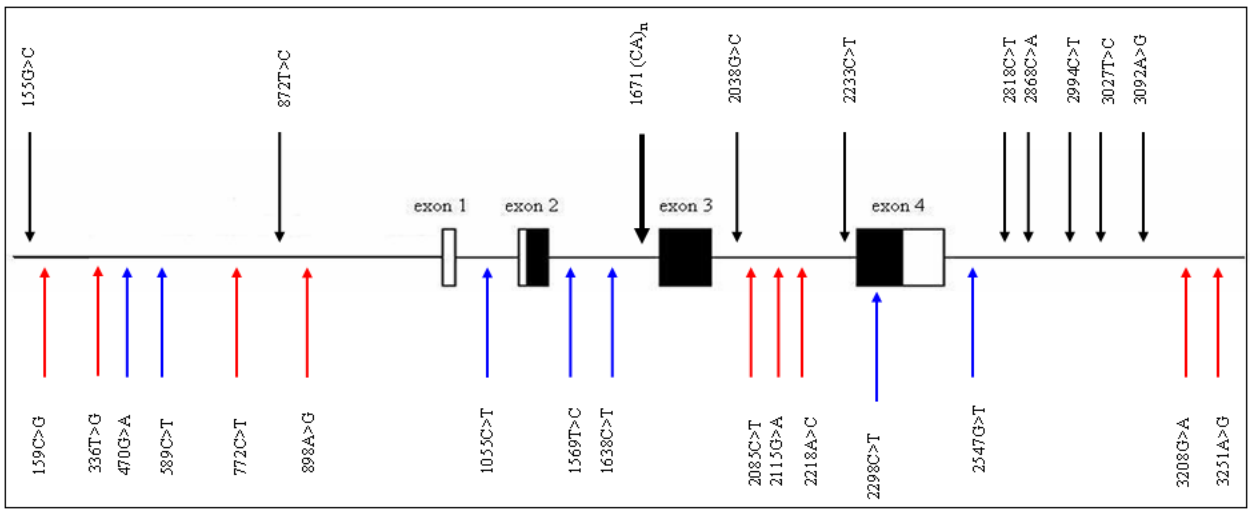


Figure 4. Schematic representation of the *APOA2* gene, showing the location of variants identified in this study. Black arrows represent variants found in both NHWs and Blacks; red arrows represent variants found in only NHWs; blue arrows represent variants found in only Blacks. Figure adapted from Martin Campos et al., 2004.

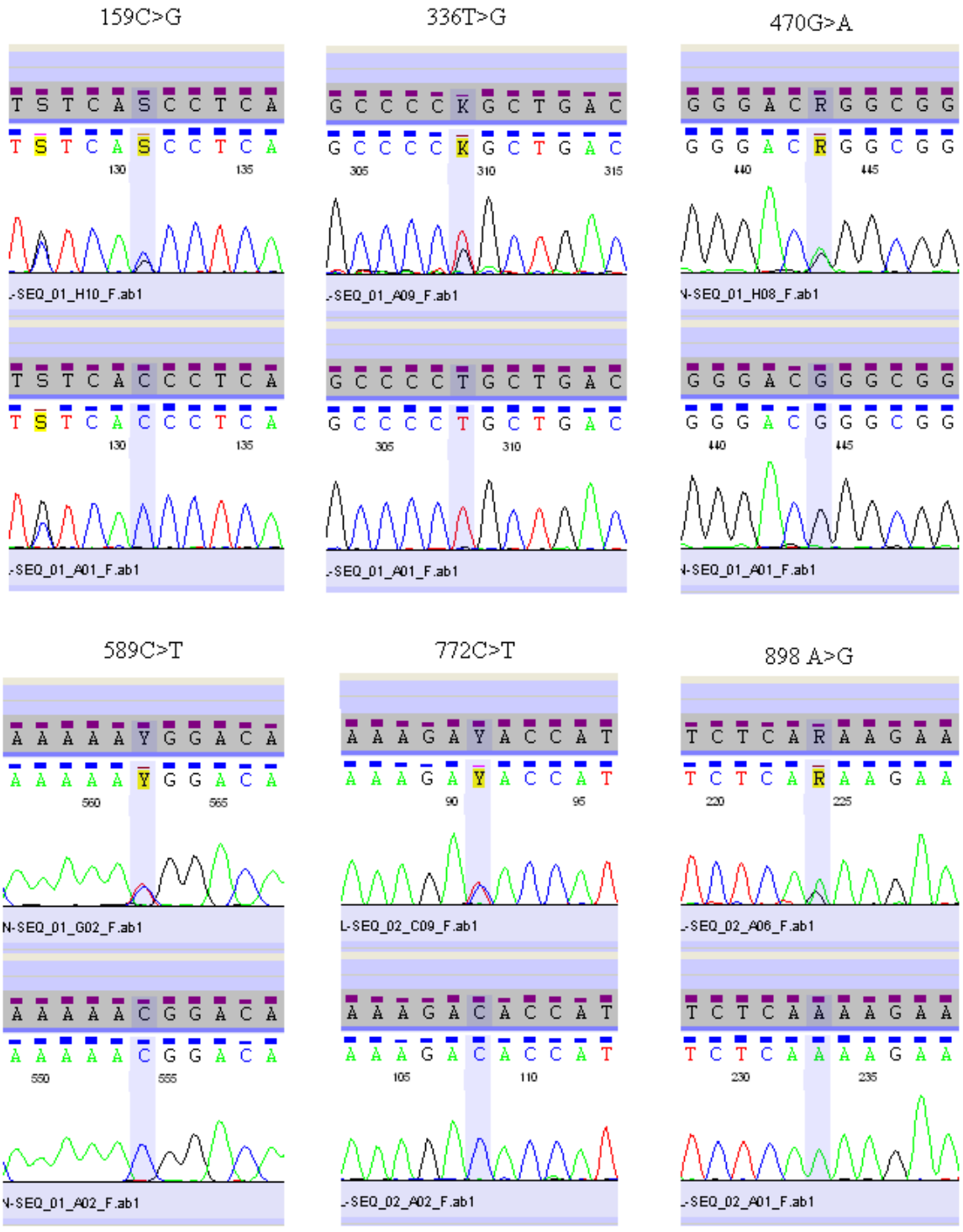


Figure 5. Chromatograms for variants at positions 159, 336, 470, 589, 772, and 898.

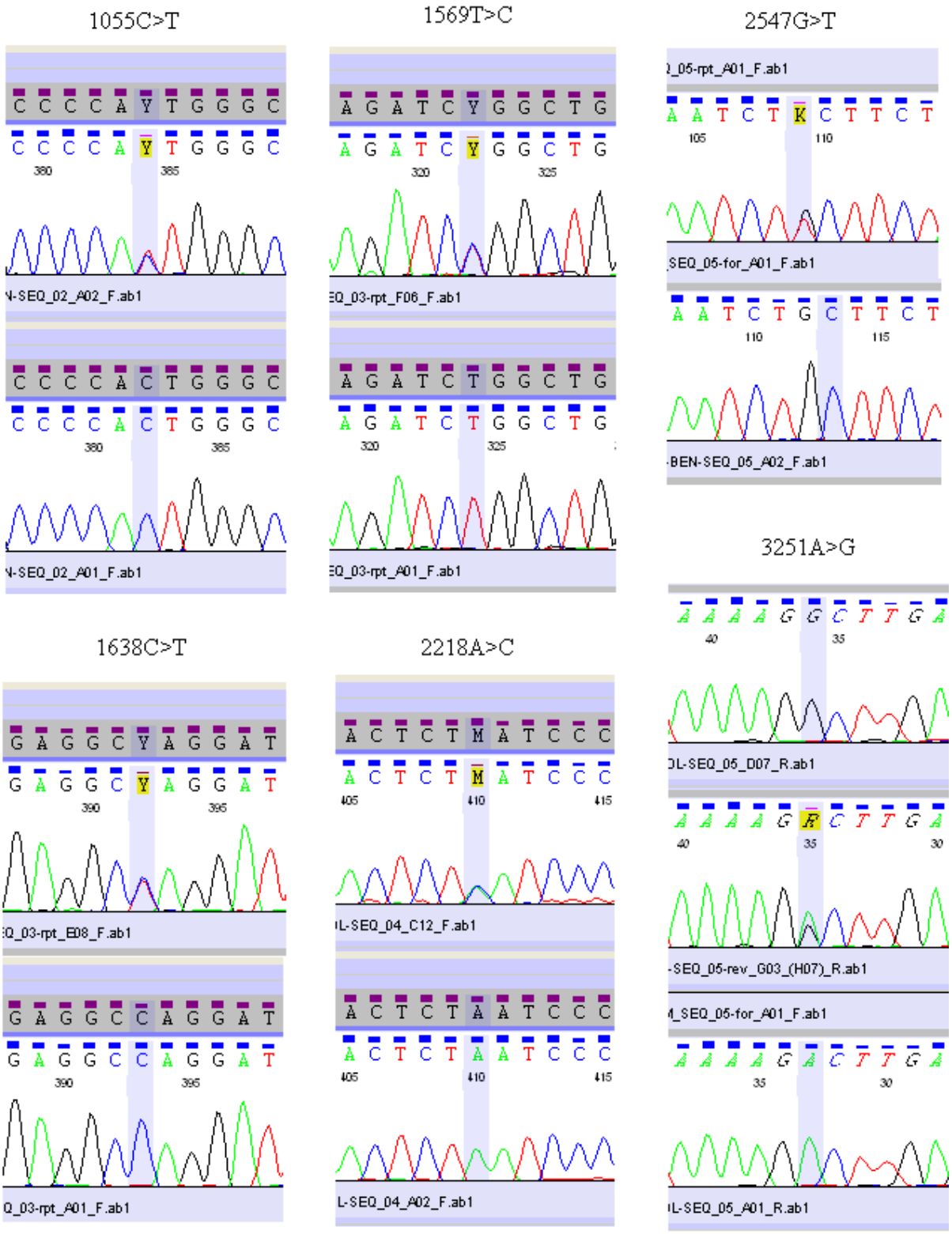


Figure 6. Chromatograms for variants at positions 1055, 1569, 2547, 1638, 2218, and 3251.

3.2 GENOTYPE DISTRIBUTION OF *APOA2* VARIANTS BETWEEN HIGH- AND LOW-HDL GROUPS

3.2.1 American Non-Hispanic Whites

Out of 18 variants (excluding the microsatellite) present in NHWs, 9 had MAF of >5%; 5 had MAF between 1% and 5%; and 4 had a MAF of less than 1%. Six out of 18 variants were newly identified, and of those, one had 5% MAF and the rest had less than 5% MAF. Table 11 and Table 12 compare the high- and low-HDL groups for variants present in heterozygous and/or homozygous states and for their genotype distribution and allele frequencies, respectively. We found a significant difference of allele frequencies for 2233C>T/rs6413453 and 2868C>A/rs12721036 SNPs between the high- and low- HDL groups (p=0.038 for each). Those two SNPs were later found to be closely linked to each other (see LD analysis section).

In NHWs, we observed 2 rare variants (excluding unconfirmed 336T>G) only in the low HDL group (898A>G and 2085C>T), and 1 rare variant only in the high HDL group (2218A>C). For 6 of 14 SNPs found in both groups, homozygosity was observed only in the low HDL group (3251A>G, 2038G>C, 2233C>T, 2818C>T, 2868C>A, and 2994C>T), although heterozygotes were seen in both high and low HDL groups. There was no variant that was observed in homozygosity in only the high HDL group.

Of the 48 individuals with low plasma levels of HDL, 7 (14.6%) had either rare variants not present in the high HDL group (N=3) or carried between 1 and 4 variants that were observed in the homozygous state only in the low HDL group (N=4). In contrast, of the 47 individuals with high plasma levels of HDL, only 1 subject (2.1%) had a rare variant not present in the low HDL group, and none carried variants that were homozygous only in the high HDL group.

Table 11. Distribution of *APOA2* variants in high and low HDL groups in NHWs. "Het" and "Hom" represent individuals heterozygous and homozygous for the minor allele, respectively.

Non-Hispanic Whites				
Variants	High HDL (*n* =47)		Low HDL (*n*= 48)	
	Het	Hom	Het	Hom
rare (MAF<5%)				
159C>G	+	-	+	-
336T>G	-	-	+	-
772C>T	+	-	+	-
898A>G	-	-	+	-
2085C>T	-	-	+	-
2115G>A	+	-	+	-
2218A>C	+	-	-	-
3208G>A	+	-	+	-
3251A>G	+	-	+	+
common (MAF≥5%)				
155G>C	+	+	+	+
872T>C	+	+	+	+
2038G>C	+	-	+	+
2233C>T	+	-	+	+
2818C>T	+	-	+	+
2868C>A	+	-	+	+
2994C>T	+	-	+	+
3027T>C	+	+	+	+
3092A>G	+	+	+	+

Table 12. Genotype and Allele Frequencies of *APOA2* Variants in High and Low-HDL Groups in NHWs.

Position 155, rs3829793						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
GG	25	(53.19)	19	(39.58)	44	46.32
GC	18	(38.30)	23	(47.92)	41	43.16
CC	4	(8.51)	6	(12.50)	10	10.53
	47		48		95	
		p =0.447				
G	0.723		0.635		0.679	
C	0.277		0.365		0.321	
		p =0.192				

Position 159, new variant						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
CC	45	(97.83)	47	(97.92)	92	97.87
CG	1	(2.17)	1	(2.08)	2	2.13
GG	0	(0.00)	0	(0.00)	0	0.00
	46		48		94	
		p =1.000				
C	0.989		0.990		0.989	
GG	0.011		0.010		0.011	
		p =0.976				

Position 336, new variant						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
TT	47	(100.00)	47	(97.92)	94	98.95
TG	0	(0.00)	1	(2.08)	1	1.05
GG	0	(0.00)	0	(0.00)	0	0.00
	47		48		95	
		p =1.000				
T	1.000		0.990		0.995	
G	0.000		0.010		0.005	
		p =0.315				

Position 772, new variant						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
CC	46	(97.87)	46	(95.83)	92	96.84
CT	1	(2.13)	2	(4.17)	3	3.16
TT	0	(0.00)	0	(0.00)	0	0.00
	47		48		95	
		p =1.000				
C	0.989		0.979		0.984	
T	0.011		0.021		0.016	
		p =0.571				

Position 872 rs5082						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
TT	15	(32.61)	16	(33.33)	31	32.98
TC	25	(54.35)	24	(50.00)	49	52.13
CC	6	(13.04)	8	(16.67)	14	14.89
	46		48		94	
		p =0.885				
T	0.598		0.583		0.590	
C	0.402		0.417		0.410	
		p =0.746				

Position 898, new variant						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
AA	47	(100.00)	47	(97.92)	94	98.95
AG	0	(0.00)	1	(2.08)	1	1.05
GG	0	(0.00)	0	(0.00)	0	0.00
	47		48		95	
		p =1.000				
A	1.000		0.990		0.995	
G	0.000		0.010		0.005	
		p =0.315				

Position 2038, rs5085						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
GG	33	(70.21)	33	(68.75)	66	69.47
GC	14	(29.79)	12	(25.00)	26	27.37
CC	0	(0.00)	3	(6.25)	3	3.16
	47		48		95	
		p =0.299				
G	0.851		0.813		0.832	
C	0.149		0.188		0.168	
		p =0.477				

Position 2085, rs5087						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
CC	47	(100.00)	47	(97.92)	94	98.95
CT	0	(0.00)	1	(2.08)	1	1.05
TT	0	(0.00)	0	(0.00)	0	0.00
	47		48		95	
		p =1.000				
C	1.000		0.990		0.995	
T	0.000		0.010		0.005	
		p =0.315				

Table 12. (Continued)

Position 2115, rs6413452						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
GG	45	(95.74)	46	(97.87)	91	96.81
GA	2	(4.26)	1	(2.13)	3	3.19
AA	0	(0.00)	0	(0.00)	0	0.00
	47		47		94	
		p =1.000				
G	0.979		0.989		0.984	
A	0.021		0.011		0.016	
		p =0.560				

Position 2218, new variant						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
AA	46	(97.87)	48	(100.00)	94	98.95
AC	1	(2.13)	0	(0.00)	1	1.05
CC	0	(0.00)	0	(0.00)	0	0.00
	47		48		95	
		p =0.495				
A	0.989		1.000		0.995	
C	0.011		0.000		0.005	
		p =0.315				

Position 2233, rs6413453						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
CC	43	(91.49)	37	(77.08)	80	84.21
CT	4	(8.51)	10	(20.83)	14	14.74
TT	0	(0.00)	1	(2.08)	1	1.05
	47		48		95	
		p =0.116				
C	0.957		0.875		0.916	
T	0.043		0.125		0.084	
		p =0.038				

Position 2818, rs12143241						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
CC	33	(70.21)	33	(68.75)	66	69.47
CT	14	(29.79)	12	(25.00)	26	27.37
TT	0	(0.00)	3	(6.25)	3	3.16
	47		48		95	
		p =0.299				
C	0.851		0.813		0.832	
T	0.149		0.188		0.168	
		p =0.477				

Position 2868, rs12721036						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
CC	43	(91.49)	37	(77.08)	80	84.21
CA	4	(8.51)	10	(20.83)	14	14.74
AA	0	(0.00)	1	(2.08)	1	1.05
	47		48		95	
		p =0.116				
C	0.957		0.875		0.916	
A	0.043		0.125		0.084	
		p =0.038				

Position 2994, rs12143180						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
CC	34	(72.34)	33	(68.75)	67	70.53
CT	13	(27.66)	12	(25.00)	25	26.32
TT	0	(0.00)	3	(6.25)	3	3.16
	47		48		95	
		p =0.364				
C	0.862		0.813		0.837	
T	0.138		0.188		0.163	
		p =0.357				

Position 3027, rs685						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
TT	34	(72.34)	28	(58.33)	62	65.26
TC	11	(23.40)	17	(35.42)	28	29.47
CC	2	(4.26)	3	(6.25)	5	5.26
	47		48		95	
		p =0.368				
T	0.840		0.760		0.800	
C	0.160		0.240		0.200	
		p =0.165				

Position 3092, rs17244530						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
AA	32	(71.11)	28	(59.57)	60	65.22
AG	11	(24.44)	16	(34.04)	27	29.35
GG	2	(4.44)	3	(6.38)	5	5.43
	45		47		92	
		p =0.554				
A	0.833		0.766		0.799	
G	0.167		0.234		0.201	
		p =0.251				

Table 12. (Continued)

Position 3208, rs12721035							Position 3251, new variant						
	High HDL		Low HDL		TOTAL			High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)		*n*	(%)	*n*	(%)	*n*	(%)
GG	43	(95.56)	43	(89.58)	86	92.47	**AA**	41	(91.11)	44	(91.67)	85	91.40
GA	2	(4.44)	5	(10.42)	7	7.53	**AG**	4	(8.89)	3	(6.25)	7	7.53
AA	0	(0.00)	0	(0.00)	0	0.00	**GG**	0	(0.00)	1	(2.08)	1	1.08
	45		**48**		**93**			**45**		**48**		**93**	
			p =0.437							*p* =0.851			
G	0.978		0.948		0.962		**A**	0.956		0.948		0.952	
A	0.022		0.052		0.038		**G**	0.044		0.052		0.048	
			p =0.277							*p* =0.808			

3.2.2 African Blacks

Of 16 variants present in Blacks (excluding the microsatellite), 7 had MAF>5%; 4 had MAF between 1% and 5%; and 5 had a MAF<1%. Six of the 16 variants were newly identified and of those, one had a MAF of 4% and the rest had less than 4% MAF. Table 13 and Table 14 compare the high- and low-HDL groups for variants present in heterozygous and/or homozygous states and for their genotype distribution and allele frequencies, respectively. We found no significant difference of allele frequencies of *APOA2* variants between the high- and low- HDL groups.

In Blacks, 3 rare variants were observed only in the low HDL group (2233C>T, 2547G>T, 2868C>A) and 2 rare variants were observed only in the high HDL group (470G>A and 1569T>C). Eleven SNPs were present in both high HDL and low HDL groups, but homozygosity was observed only in the low HDL group for 6 variants (155G>C, 2038G>C, 2818C>T, 2994C>T, 3027T>C, 3092A>G) and only in the high HDL group for 1 variant (872T>C).

Of the 47 individuals with low plasma levels of HDL, 4 (8.5%) had either rare variants not present in the high HDL group (N=3) or carried 5 variants that were observed in the homozygous state only in the low HDL group (N=1). Of the 48 individuals with high plasma levels of HDL, 3 (6.3%) had either a rare variant not present in the low HDL group (N=2), or carried a variant that was homozygous only in the high HDL group (N=1).

Table 13. Distribution of *APOA2* variants in high and low HDL groups in African Blacks. "Het" and "Hom" represent individuals heterozygous and homozygous for the minor allele, respectively.

Blacks				
Variants	High HDL (*n* =48)		Low HDL (*n* =47)	
	Het	Hom	Het	Hom
rare (MAF<5%)				
470G>A	+	-	-	-
589C>T	+	-	+	-
1055C>T	+	-	+	-
1569T>C	+	-	-	-
1638C>T	+	-	+	-
2233C>T	-	-	+	-
2298C>T	+	-	+	-
2547G>T	-	-	+	-
2868C>A	-	-	+	-
common (MAF≥5%)				
155G>C	+	-	+	+
872T>C	+	+	+	-
2038G>C	+	-	+	+
2818C>T	+	-	+	+
2994C>T	+	-	+	+
3027T>C	+	-	+	+
3092A>G	+	-	+	+

Table 14. Genotype and Allele Frequencies of *APOA2* Variants in High and Low-HDL Groups in Blacks.

Position 155, rs3829793						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
GG	41	(85.42)	43	(91.49)	84	88.42
GC	7	(14.58)	3	(6.38)	10	10.53
CC	0	(0.00)	1	(2.13)	1	1.05
	48		47		95	
		p =0.317				
G	0.927		0.947		0.937	
C	0.073		0.053		0.063	
		p= 0.575				

Position 470, new variant						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
GG	47	(97.92)	47	(100.00)	94	98.95
GA	1	(2.08)	0	(0.00)	1	1.05
AA	0	(0.00)	0	(0.00)	0	0.00
	48		47		95	
		p =1.000				
G	0.990		1.000		0.995	
A	0.010		0.000		0.005	
		p= 0.315				

Position 589, new variant						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
CC	47	(97.92)	44	(93.62)	91	95.79
CT	1	(2.08)	3	(6.38)	4	4.21
TT	0	(0.00)	0	(0.00)	0	0.00
	48		47		95	
		p =0.362				
C	0.990		0.968		0.979	
T	0.010		0.032		0.021	
		p= 0.303				

Position 872, new variant						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
TT	31	(64.58)	28	(59.57)	59	62.11
TC	16	(33.33)	19	(40.43)	35	36.84
CC	1	(2.08)	0	(0.00)	1	1.05
	48		47		95	
		p =0.597				
T	0.813		0.798		0.805	
C	0.188		0.202		0.195	
		p= 0.799				

Position 1055, new variant						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
CC	44	(91.67)	43	(91.49)	87	91.58
CT	4	(8.33)	4	(8.51)	8	8.42
TT	0	(0.00)	0	(0.00)	0	0.00
	48		47		95	
		p =1.000				
C	0.958		0.957		0.958	
T	0.042		0.043		0.042	
		p= 0.976				

Position 1569, new variant						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
TT	46	(97.87)	47	(100.00)	93	98.94
TC	1	(2.13)	0	(0.00)	1	1.06
CC	0	(0.00)	0	(0.00)	0	0.00
	47		47		94	
		p =1.000				
T	0.989		1.000		0.995	
C	0.011		0.000		0.005	
		p= 0.315				

Position 1638, new variant						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
CC	44	(93.62)	45	(95.74)	89	94.68
CT	3	(6.38)	2	(4.26)	5	5.32
TT	0	(0.00)	0	(0.00)	0	0.00
	47		47		94	
		p =1.000				
C	0.968		0.979		0.973	
T	0.032		0.021		0.027	
		p= 0.650				

Position 2038, rs5085						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
GG	38	(79.17)	40	(85.11)	78	82.11
GC	10	(20.83)	6	(12.77)	16	16.84
CC	0	(0.00)	1	(2.13)	1	1.05
	48		47		95	
		p =0.412				
G	0.896		0.915		0.905	
C	0.104		0.085		0.095	
		p=0.653				

Table 14. (Continued)

Position 2233, rs6413453						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
CC	48	(100.00)	46	(97.87)	94	98.95
CT	0	(0.00)	1	(2.13)	1	1.05
TT	0	(0.00)	0	(0.00)	0	0.00
	48		**47**		**95**	
		p =0.495				
C	1.000		0.989		0.995	
T	0.000		0.011		0.005	
		p= 0.315				

Position 2298, rs5088						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
CC	45	(93.75)	45	(95.74)	90	94.74
CT	3	(6.25)	2	(4.26)	5	5.26
TT	0	(0.00)	0	(0.00)	0	0.00
	48		**47**		**95**	
		p =1.000				
C	0.969		0.979		0.974	
T	0.031		0.021		0.026	
		p= 0.667				

Position 2547, new variant						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
GG	48	(100.00)	46	(97.87)	94	98.95
GT	0	(0.00)	1	(2.13)	1	1.05
TT	0	(0.00)	0	(0.00)	0	0.00
	48		**47**		**95**	
		p =0.495				
G	1.000		0.989		0.995	
T	0.000		0.011		0.005	
		p= 0.315				

Position 2818, rs12143241						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
CC	38	(79.17)	40	(85.11)	78	82.11
CT	10	(20.83)	6	(12.77)	16	16.84
TT	0	(0.00)	1	(2.13)	1	1.05
	48		**47**		**95**	
		p =0.412				
C	0.896		0.915		0.905	
T	0.104		0.085		0.095	
		p= 0.653				

Position 2868, rs12721036						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
CC	48	(100.00)	46	(97.87)	94	98.95
CA	0	(0.00)	1	(2.13)	1	1.05
AA	0	(0.00)	0	(0.00)	0	0.00
	48		**47**		**95**	
		p =0.495				
C	1.000		0.989		0.995	
A	0.000	0.315	0.011		0.005	
		p= 0.315				

Position 2994, rs12143180						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
CC	42	(87.50)	44	(93.62)	86	90.53
CT	6	(12.50)	2	(4.26)	8	8.42
TT	0	(0.00)	1	(2.13)	1	1.05
	48		**47**		**95**	
		p =0.268				
C	0.938		0.957		0.947	
T	0.063		0.043		0.053	
		p= 0.537				

Position 3027, rs685						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
TT	38	(79.17)	40	(85.11)	78	82.11
TC	10	(20.83)	6	(12.77)	16	16.84
CC	0	(0.00)	1	(2.13)	1	1.05
	48		**47**		**95**	
		p =0.412				
T	0.896		0.915		0.905	
C	0.104		0.085		0.095	
		p= 0.653				

Position 3092, rs17244530						
	High HDL		Low HDL		TOTAL	
	n	(%)	*n*	(%)	*n*	(%)
AA	38	(79.17)	40	(85.11)	78	82.11
AG	10	(20.83)	6	(12.77)	16	16.84
GG	0	(0.00)	1	(2.13)	1	1.05
	48		**47**		**95**	
		p =0.413				
A	0.896		0.915		0.905	
G	0.104		0.085		0.095	
		p= 0.653				

3.3 LD AND TAGGER ANALYSES OF *APOA2* VARIANTS USING HAPLOVIEW

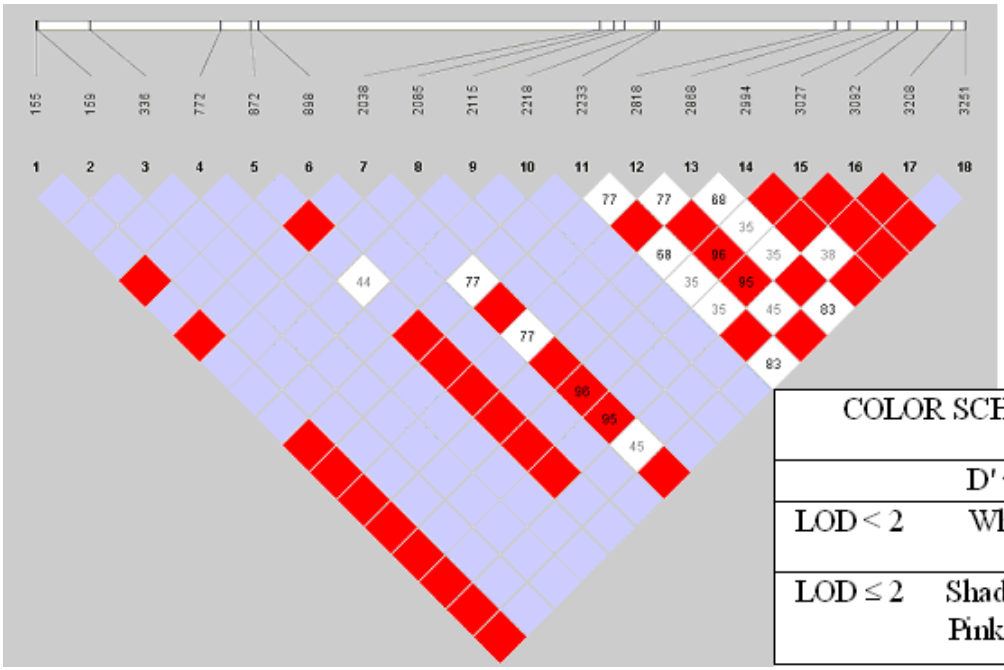
SNPs that are located close to one another are often inherited together and are compiled into haplotypes. "TagSNPs" are a group of SNPs that uniquely identify those haplotypes. Identification of TagSNPs using Tagger, helps to reduce the number of common SNPs to be screened to gain information.

3.3.1 American Non-Hispanic Whites

LD and Tagger analysis (using Haploview) of 18 variants in NHWs (excluding the microsatellite) using an r^2 cutoff of 0.95 identified 14 bins. Table 15 summarizes the Tagger results, and Figure 6 shows the LD plot. As expected, all the rare variants were in individual bins and had to be screened individually.

Table 15. Tagger Results. The variants that are marked in bold are considered for genotyping the entire population.

BIN	Location of Variants Captured	rs# of the selected variants
1	2038, 2818, **2994**	rs12143180
2	**3027**, 3092	rs685
3	**2233**, 2868	rs6413453
4	**772**	-
5	**2218**	-
6	**898**	-
7	**336**	-
8	**155**	rs3829793
9	**3208**	rs12721035
10	**3251**	-
11	**2085**	rs5087
12	**159**	-
13	**2115**	rs6413452
14	**872**	rs5082



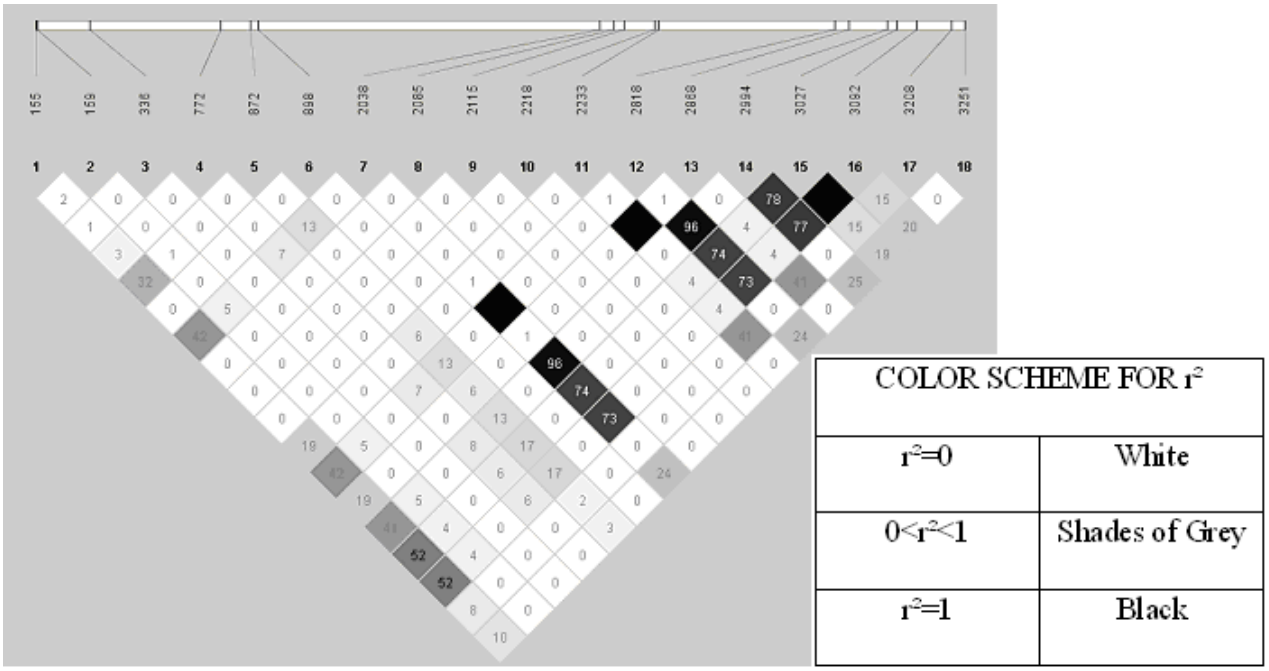


Figure 7. LD Analysis for NHWs

Ten common SNPs (MAF≥5%) were observed in 6 bins. Therefore it would be sufficient to screen only 6 variants to gain information on all 10 common variants. TaqMan assays (pre-made or custom) were available for at least one SNP from 5 out of 6 bins of common SNPs (rs12143180, rs6413453, rs3828793, rs5082, 3251). Custom TaqMan design failed for one of the common SNP bins. The remaining 8 non-common variants with MAF<5% (that were only present in NHWs) were first evaluated for presence of a TaqMan assay (3 were available), followed by evaluation by RFLP when TaqMan assay was not available.

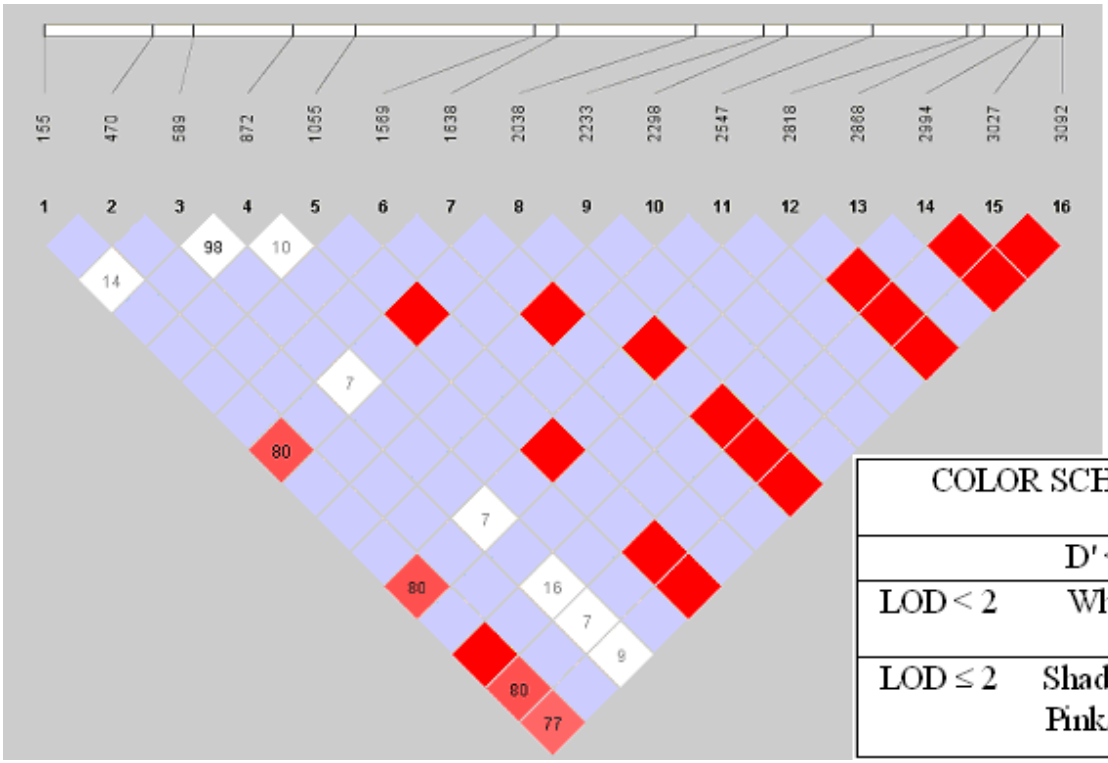
3.3.2 African Blacks

LD and Tagger analysis (using Haploview) of 16 variants in Blacks (excluding the microsatellite) using an r^2 cutoff of 0.95 identified 12 bins. Table 16 summarizes the Tagger results, and Figure 7 shows the LD plot.

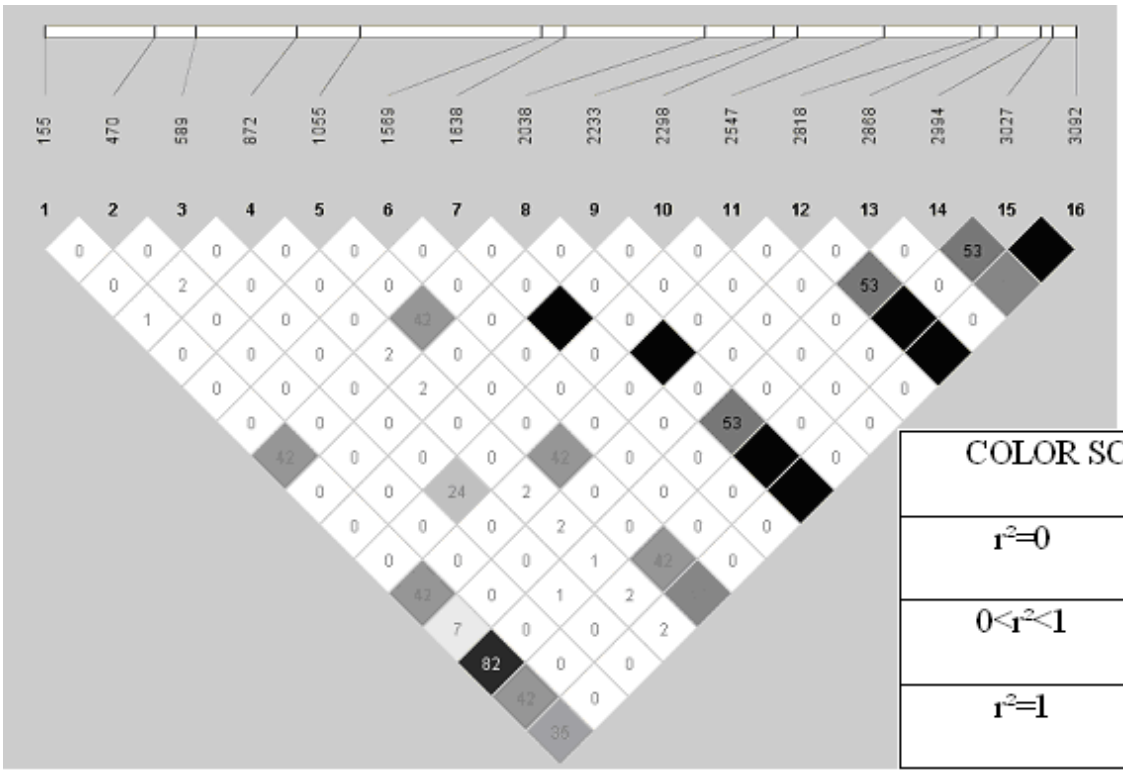
Table 16. Tagger Results. The variants that are marked in bold are considered for genotyping the entire population.

BIN	Variants Captured	rs# of the selected variants
1	2038, 2818, **3027**, 3092	rs685
2	1638, **2298**	rs5088
3	**2868**	rs12721036
4	**470**	-
5	**2994**	rs12143180
6	**589**	-
7	**155**	rs3829793
8	**2547**	-
9	**872**	rs5082
10	**1055**	-
11	**2233**	rs6413453
12	**1569**	-

Seven common SNPs (MAF≥5%) were presented in 4 bins. Therefore, it would be sufficient to screen only 4 variants to gain information about all 7 common variants. TaqMan assays (pre-made or custom) were available for at least 3 SNPs from 3 out of 4 bins of common SNPs (rs12143180, rs3829793, rs5082). Custom TaqMan design failed for one of the common SNP bins. The remaining 9 non-common variants less than 5% MAF (7 of 9 were only present in Blacks) were first evaluated for presence of a TaqMan assay (2 were available), followed by evaluation by RFLP when TaqMan assay was not available.



COLOR SCHEME FOR D'		
	D' < 1	D' = 1
LOD < 2	White	Blue
LOD ≤ 2	Shades of Pink/Red	Bright Red



COLOR SCHEME FOR r^2	
$r^2=0$	White
$0<r^2<1$	Shades of Grey
$r^2=1$	Black

Figure 8. LD analysis of African Blacks

3.4 GENOTYPING SCREENING IN NHWs AND BLACKS

3.4.1 RFLP

3.4.1.1 American Non-Hispanic Whites

Among the 5 rare variants in NHWs, for which there was no available TaqMan assay, 2 did not affect any restriction site. For 2 of the 3 remaining variants, we were able to successfully design RFLP PCR primers (variants at position 336 and 159).

3.4.1.2 African Blacks

Among the 6 rare variants in Blacks, for which there was no available TaqMan assay, 2 did not affect any restriction site. For 2 of the remaining 4 variants, we were able to successfully design RFLP PCR primers (variants at positions 1055 and 1569).

3.4.1.3 NHWs and Blacks

Our RFLP analysis confirmed the variants at locations 159, 1055, and 1569 by using all known heterozygous individuals from 159 and 1569 (N=2 and N=1, respectively) and by using 3 of 8 known heterozygous individuals from 1055. We failed to confirm the variation identified (by sequencing) at location 336 using RFLP. Since the only known heterozygote showed the same pattern as the known homozygous wild-types, we concluded that this was not a genuine variant, but sequencing noise. The RFLP results are shown in Figure 8, Figure 9, Figure 10 and Figure 11.

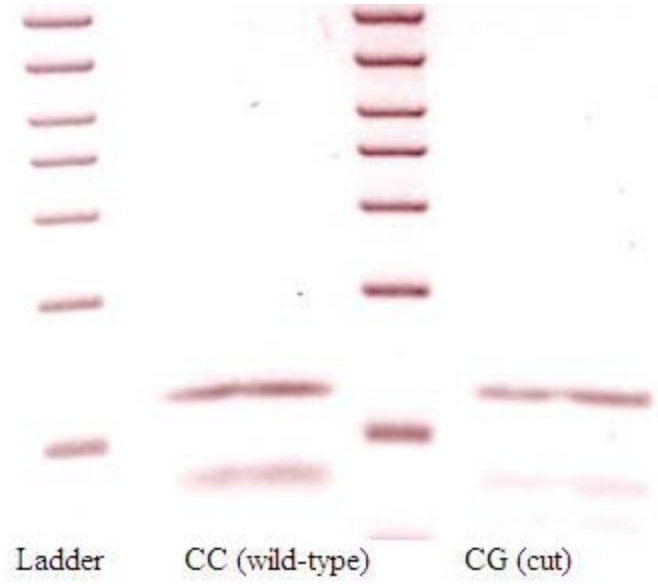


Figure 9. Variant location 159.

Gel picture of the Restriction Fragment Length Polymorphism (RFLP) analysis of PCR-amplified DNA samples from one wild-type CC homozygous individual and one CG heterozygous individual (as identified by sequencing). The PCR fragment has both a polymorphic and non-polymorphic restriction site, so we observe digestion even in wild-type homozygotes. *Cvi*KI-1 digestion generates two fragments of 129 and 85 base pairs in the CC individual and four fragments of 129, 85, 65, and 20 in the CG individual (the smallest 20 bp fragment cannot be visualized with our electrophoresis conditions). The third band (65 base pairs) in the CG individual is faint, but can still be distinguished.

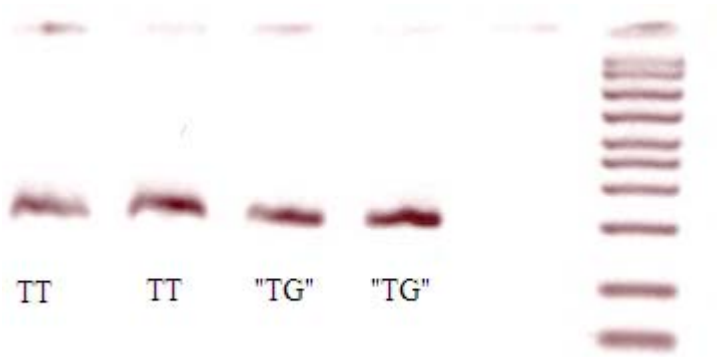


Figure 10. Variant location 336.

Gel picture of the RFLP analysis of PCR-amplified DNA samples from two wild-type TT homozygous individuals and two TG heterozygous individuals (as identified by sequencing). The G allele at this location is expected to create a restriction site for *Msp*I, therefore the digestion generates only 1 fragment of 200bps in homozygous TT individuals vs. 2 fragments of 113bps and 87bps in a heterozygous TG individual. We were not able to observe that pattern, and all samples appeared as uncut TT individuals, suggesting that this variant may not be real.

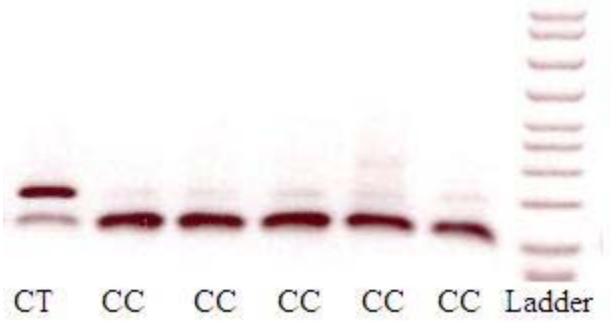


Figure 11. Variant location 1055.

Gel picture of the RFLP analysis of PCR-amplified DNA samples from one CT heterozygous and five wild-type CC homozygous individuals (as identified by sequencing.) The C>T variant abolishes a recognition site for enzyme *Bsr*-I at location 1055 in *APOA2*. Restriction endonuclease (*Bsr*I) digestion generates 2 fragments in CC homozygous individuals (99bps and 87bps fragments which are not distinguishable at this resolution). In homozygous TT individuals, the expected fragment size is 186bp (undigested), and in CT heterozygous individuals, we expect 1 uncut fragment (186bps) and 2 cut fragments (indistinguishable 99bps and 87bps).

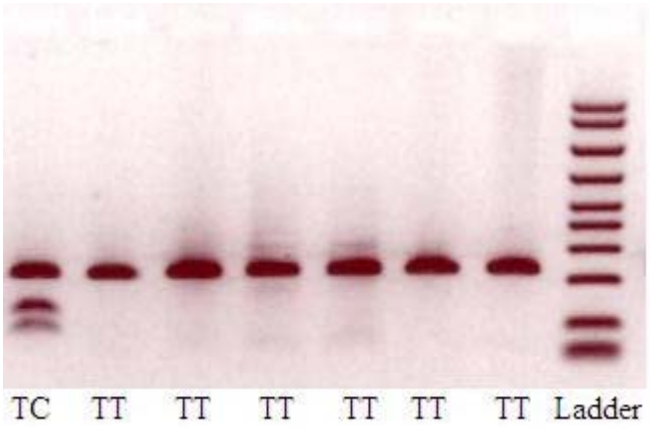


Figure 12. Variant location 1569.

Gel picture of the RFLP analysis of PCR-amplified DNA samples from one TC heterozygous and 6 wild-type TT homozygous individuals (as identified by sequencing.) The T>C variant creates a recognition site for restriction endonuclease *Msp*I at location 1569 in *APOA2*. *Msp*I digestion generates one fragment of 218 base pairs in a TT individual, and three fragments of 218, 128, and 90 base pairs in the TC individual. Expected fragments in a homozygous CC individual would be 128 and 90bps.

3.4.1.4 Pooling Strategy

To expedite the genotyping process, we employed a pooling strategy, in which 10 samples were combined after PCR amplification, before RFLP restriction digest. We tried pooling 10 samples (one of which carried a known variant) to determine whether we could detect

variants on a larger scale. A post-PCR pooling method worked for variants 1055 and 1569, but not for 159. Variant 159 was faint in single-sample analysis and could not be detected using the pooling method. We will also try 8- and 5-sample post-PCR pooling for 155 and 1569 to determine whether we can achieve stronger bands (to decrease the chance for false negatives).

The following figure (Figure 12) depicts all the variants we identified in *APOA2* relative to their location in the gene. Variant 336 was not included as we could not confirm it. The color FASTA that is used in this figure was obtained from Seattle SNPs Database and revised according to the variants identified by sequencing in this study. The variants identified in public databases are shown in **black font**, and the new variants identified in this study are shown in **red font**. The variants that are observed only in one heterozygous individual, and were not confirmed by either RFLP or TaqMan are highlighted in yellow, and remain to be confirmed. The remaining variants were either observed in more than one individual and/or confirmed with RFLP or TaqMan. The minor alleles and their frequencies are depicted on the right (MAF in NHWs, MAF in Blacks). When only one frequency is given, the variant was only observed in one population.

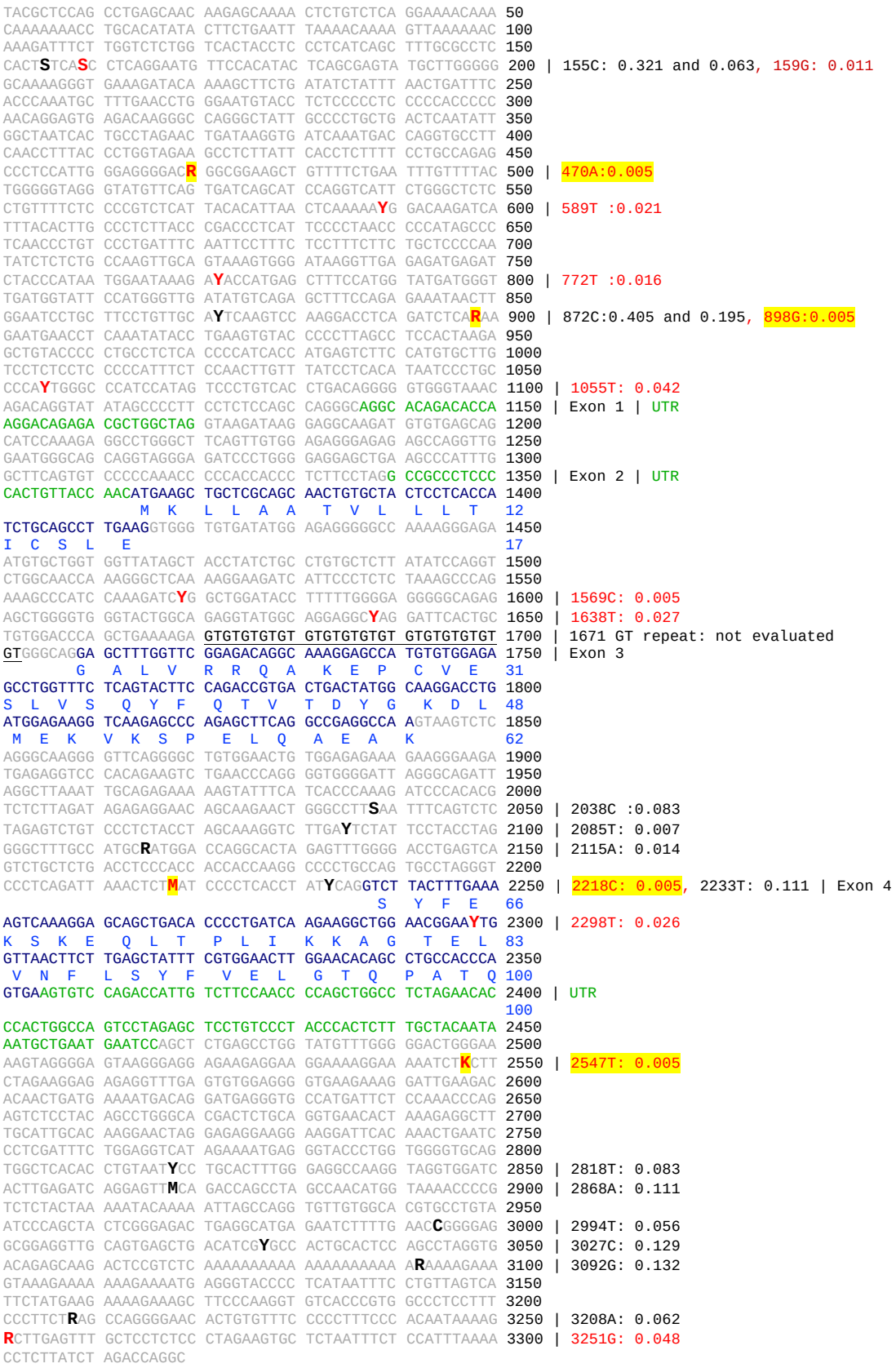


Figure 13. *APOA2* sequence variants identified in our study

3.4.2 TaqMan SNP Genotyping

We have genotyped 5 variants in Blacks (rs12143180, rs3829793, rs5088, rs6413453, rs5082) and 8 variants in NHWs (rs12143180, rs6413453, rs3828793, rs6413452, rs12721035, rs5082, rs5087, 3251) using either pre-made or custom-made TaqMan SNP Genotyping Assays. Genotyping call rates for a total of 9 assays (4 common to both populations) are shown in Table 17. The genotyping discrepancy rate was less than 1.5% for each variant based on 10% repeat of the samples.

Table 17. Genotyping Call Rates for TaqMan (given in percentages)

Location	NHWs	Blacks
rs12143180	98.88	96.45
rs3829793	98.88	98.98
rs6413453	99.84	98.48
rs5082	99.2	98.73
rs5088	-	98.73
3251	99.84	-
rs5087	99.52	-
rs6413452	99.84	-
rs12721035	99.84	-

Comparison of the allele frequencies and genotype distribution of the screened variants between NHWs and Blacks is given in Table 18. None of the variants showed any deviation from HWE ($p>0.05$) in each race group. All the variants that were present in both populations showed significant differences in genotype and allele frequency when compared between races ($p<0.0001$).

Table 18. Genotype and Allele frequencies of *APOA2* variants genotyped with TaqMan assays. The statistical comparison between NHWs and Blacks was made when the variant was present in both populations. NA= not applicable.

APOA2 variant	Genotype/Allele	NHWs	Blacks	APOA2 variant	Genotype/Allele	NHWs	Blacks
rs6413453	*Genotypes*	*n*(%)	*n*(%)	rs5088	*Genotypes*	*n*(%)	*n*(%)
	CC	481(77.21)	765(98.58)		CC	-	741(95.49)
	CT	138(22.15)	11(1.42)		CT	-	34(4.12)
	TT	4(0.64)	0		TT	-	3(0.39)
	total	623	776		total	-	776
		$p<0.0001$				p=NA	
	Alleles				*Alleles*		
	C	0.883	0.993		C	-	0.973
	T	0.117	0.007		T	-	0.026
		$p<0.0001$				p=NA	

APOA2 variant	Genotype/Allele	NHWs	Blacks	APOA2 variant	Genotype/ Allele	NHWs	Blacks
rs3829793	*Genotypes*	*n*(%)	*n*(%)	rs12143180	*Genotypes*	*n*(%)	*n*(%)
	GG	255(41.32)	650(83.33)		CC	419(67.91)	652(85.79)
	GC	279(45.21)	123(15.77)		CT	177(28.69)	104(13.68)
	CC	83(13.45)	7(0.90)		TT	21(3.40)	4(0.53)
	total	617	780		total	617	760
		$p<0.0001$				$p<0.0001$	
	Alleles				*Alleles*		
	G	0.639	0.912		C	0.883	0.926
	C	0.361	0.088		T	0.118	0.074
		$p<0.0001$				$p<0.0001$	

APOA2 variant	Genotype/Allele	NHWs	Blacks	APOA2 variant	Genotype/Allele	NHWs	Blacks
rs6413452	*Genotypes*	*n*(%)	*n*(%)	rs5087	*Genotypes*	*n*(%)	*n*(%)
	GG	609(97.75)	-		CC	619(99.68)	-
	GA	14(2.25)	-		CT	2(0.32)	-
	AA	0	-		TT	0	-
	total	623	-		total	621	-
		p=NA				p=NA	
	Alleles				*Alleles*		
	G	0.989	-		C	0.998	-
	A	0.011	-		T	0.002	-
		p=NA				p=NA	

Table 18 Continued

APOA2 variant	Genotype/Allele	NHWs	Blacks
rs12721035	*Genotypes*	*n*(%)	*n*(%)
	GG	549(88.12)	-
	GA	74(11.88)	-
	AA	0	-
	total	623	-
		p=NA	
	Alleles		
	G	0.941	-
	A	0.059	-
		p=NA	

APOA2 variant	Genotype/Allele	NHWs	Blacks
3251A>G	*Genotypes*	*n*(%)	*n*(%)
	AA	580(93.10)	-
	AG	40(6.42)	-
	GG	3(0.48)	-
	total	623	-
		p=NA	
	Alleles		
	A	0.963	-
	G	0.037	-
		p=NA	

APOA2 variant	Genotype/Allele	NHWs	Blacks
rs5082	*Genotypes*	*n*(%)	*n*(%)
	TT	232 (37.54)	507(65.63)
	TC	297(47.90)	242(30.61)
	CC	90(14.56)	29(3.76)
	total	618	771
		p<0.0001	
	Alleles		
	T	0.615	0.809
	C	0.385	0.191
		p<0.0001	

3.5 LD ANALYSIS USING GENOTYPE DATA FROM ENTIRE POPULATION

When we repeated LD analysis using data from the entire sample, we observed similar LD patterns as were first seen in the smaller sample. Figure 13 and Figure 14 show the *D'* and r^2 analysis.

3.5.1 Non-Hispanic Whites

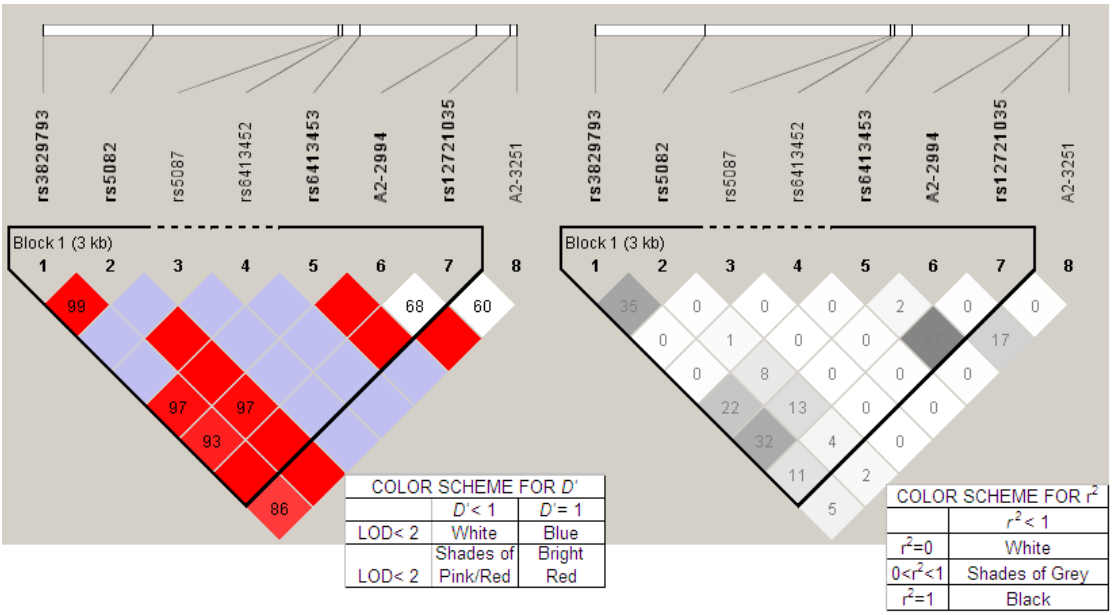


Figure 14. *D'* and r^2 analyses in NHWs

3.5.2 Blacks

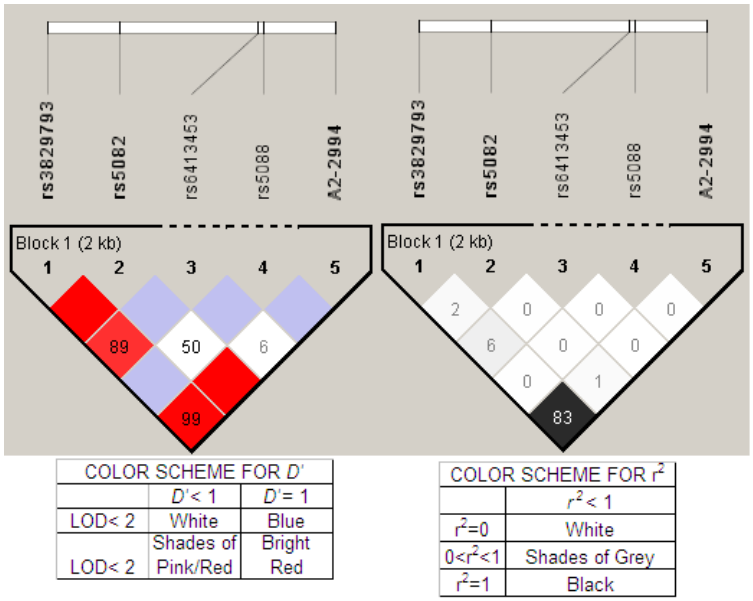


Figure 15. *D'* and r^2 analyses in Blacks

3.6 ASSOCIATION OF *APOA2* VARIANTS WITH TOTAL HDL LEVELS

We analyzed the association of *APOA2* variants (genotyped using TaqMan assays) in relation to total HDL levels (using square root transformed HDL levels) separately in males and females within each ethnic group (Table 19 for NHWs and Table 20 for Blacks). The covariates used in NHWs were BMI for males, and BMI, age, and smoking (currently) for females. The covariates used in Blacks were waist measurement and carbohydrate intake (as percentage calories) for males, and smoking (currently), waist measurement, and minutes of jogging or cycling per day for females.

3.6.1 American NHWs

No significant association was seen in males. In females, we found a significant association for 2 variants; 2233C>T/rs6413453 (p=0.028) and 3251A>G (p=0.023). The total HDL levels were significantly lower in females who were homozygous for the minor allele for each of these variants. Because there were only 3 females who were homozygous for the minor alleles of these 2 variants, we repeated our analysis in females by comparing wild-type allele homozygotes vs. minor allele heterozygotes+homozygotes. We still observed a slight significance for the association of rs6413453 (p=0.043), however there was no longer a significant association with 3251A>G (p=0.803).

Table 19. Genotype distribution, mean HDL levels, and adjusted *p*-values for 8 *APOA2* variants in NHWs

	NHW MALES			NHW FEMALES		
rs3829793	**GG (127)**	**GC (125)**	**CC (40)**	**GG (127)**	**GC (154)**	**CC (43)**
HDL-C	44.79±11.64	43.69±10.46	43.38±9.20	56.72±14.17	57.23±13.77	54.60±15.81
	p=0.87[a]			*p*=0.34[b]		
rs5082	**TT (112)**	**TC(136)**	**CC (45)**	**TT (120)**	**TC (161)**	**CC (44)**
HDL-C	43.62±10.06	44.92±11.85	42.56±9.32	57.25±15.84	57.18±13.73	54.61±13.18
	p=0.35[a]			*p*=1.00[b]		
rs5087	**CC (292)**	**CT (2)**	**TT (0)**	**CC (326)**	**CT (0)**	**TT (0)**
HDL-C	44.04±10.85	36.00±9.90	-	56.80±14.45	-	-
	p=0.21[a]			*p*= NA		
rs6413452	**GG (284)**	**GA (11)**	**AA (0)**	**GG (324)**	**GA (3)**	**AA (0)**
HDL-C	43.72±10.73	51.18±11.44	-	56.90±14.50	51.33±3.06	-
	p=0.08[a]			*p*=0.92[b]		
rs6413453	**CC (232)**	**CT (61)**	**TT (2)**	**CC (248)**	**CT (77)**	**TT (2)**
HDL-C	44.44± 11.35	42.36± 8.70	42.00±1.41	57.34± 14.54	55.60± 14.01	39.00±5.66
	p=0.47[a]			***p*=0.03**[b]		
rs12143180	**CC (196)**	**CT (85)**	**TT (11)**	**CC (222)**	**CT (92)**	**TT (10)**
HDL-C	44.17±10.80	43.92±11.46	42.55±7.88	56.95±14.31	57.42±14.92	51.20±13.83
	p=0.81[a]			*p*=0.25[b]		
rs12721035	**GG (261)**	**GA (34)**	**AA (0)**	**GG (287)**	**GA (40)**	**AA (0)**
HDL-C	44.01±10.97	43.88±9.86	-	56.76±14.54	57.28±13.84	-
	p=0.91[a]			*p*=0.56[b]		
3251A>G	**AA (272)**	**AG (20)**	**GG (2)**	**AA (307)**	**AG (20)**	**GG (1)**
HDL-C	43.69± 10.61	48.10± 13.59	40.50±3.54	56.83± 14.31	58.50± 15.09	26.00±NA
	p=0.18 [a]			***p*=0.02**[b]		

[a]*p*-values for square root-transformed HDL levels, adjusting for "BMI"
[b]*p*-values for square root-transformed HDL levels, adjusting for "age, smoking status and BMI"

3.6.2 African Blacks

None of the 5 variants showed an association with total HDL levels in males or females (Table 20).

Table 20. Genotype distribution, mean HDL levels, and adjusted *p*-values for 5 *APOA2* variants in Blacks.

	BLACK MALES			**BLACK FEMALES**		
rs12143180	**CC (402)**	**CT (60)**	**TT (1)**	**CC (236)**	**CT (39)**	**TT (2)**
HDL-C	45.76±12.96	47.58±11.39	56.9±NA	51.01±12.93	53.13±11.30	40.60±14.99
	P=0.41[a]			p=0.26[b]		
rs5088	**CC (451)**	**CT (19)**	**TT (3)**	**CC (270)**	**CT (14)**	**TT (0)**
HDL-C	46.11±12.80	44.98±10.60	51.13±5.95	51.13±12.56	50.91±14.93	-
	P=0.47[a]			p=0.98[b]		
rs3829793	**GG (399)**	**GC (72)**	**CC (3)**	**GG (236)**	**GC (46)**	**CC (3)**
HDL-C	45.84±12.96	47.28±11.15	44.03±13.72	50.68±12.78	53.69±11.72	42.70±11.21
	P=0.53[a]			p=0.16[b]		
rs6413453	**CC (468)**	**CT (3)**	**TT (0)**	**CC (276)**	**CT (8)**	**TT (0)**
HDL-C	46.10±12.72	36.00±2.03	-	51.08±12.65	52.56±13.55	-
	P=0.22[a]			p=0.90[b]		
rs5082	**TT (316)**	**TC (141)**	**CC (18)**	**TT (179)**	**TC (92)**	**CC (11)**
HDL-C	46.04±12.81	46.27±12.88	45.47±8.55	51.14±12.88	51.35±12.51	51.14±10.93
	P=1.00[a]			p=0.97[b]		

[a]*p*-values for square root-transformed HDL levels, adjusting for "waist measurement and carbohydrate intake"
[b]*p*-values for square root-transformed HDL levels, adjusting for "smoking, waist measurement, and exercise"

4.0 DISCUSSION

ApoA-II is the second major protein after apoA-I in HDL particles and plays a role in HDL metabolism and reverse cholesterol transport. The role of apoA-II in atherogenesis is controversial, but it is not a neutral molecule. Studies in mouse models have also produced varying results. The *APOA2* gene clearly requires further study due to its biological significance and lack of consensus in its biological function.

Our study was aimed to evaluate the *APOA2* gene in detail by resequencing DNA samples from healthy individuals with extreme levels of HDL (upper and lower 5th percentiles). The purpose of using this method was to detect the rare causative variants in addition to common variants allowing us to test both the "common variant-common disease" and "rare variant-common disease" hypotheses.

Joint analysis of recent genome-wide association studies have identified several genes involved in lipid metabolism (Willer et al., 2008 and Kathiresan el al., 2008). Willer et al. (2008) confirmed 11 previously implicated loci and identified several new loci by performing meta-analysis of three genome-wide association studies. Kathiresan et al. (2008) showed that common SNPs at 18 loci are reproducibly associated with lipid metabolism including 6 new loci. According to both studies, the major loci associated with HDL-cholesterol at genome-wide significance level were *ABCA1*, the *APOA1-APOC3-APOA4-APOA5* cluster, *CETP*, *LIPC*, *LIPG*, *LPL*, and *GALNT2*. While *APOA2* is not reported as one of the major loci in these studies,

the role of *APOA2* in lipid metabolism cannot be dismissed. First of all, several loci with significant impact may be overlooked in genome-wide association studies, due to the fact that a very low cutoff *p*-value is often used ($P<5 \times 10^{-7}$). Second, the goal of these genome-wide association studies is to identify common SNPs throughout the genome that are associated with lipid concentrations, testing the "common variant-common disease" hypothesis, but not being able to identify the common variants with small effects or the rare variants. When Willer et al. (2008) applied multiple regression models during their meta-analysis; they found that the variants identified in their study accounted for only ~5 to 8% of the variation in the lipid traits in their population. Because early studies have suggested high heritability rates for lipid traits (ranging from 24-83% for HDL), they concluded that additional genetic factors remain to be identified, such as common variants with small affects and rare variants.

Cohen et al. (2004) used a sequencing strategy (coding regions and splice sites) to evaluate the rare variants in 3 genes (*ABCA1*, *APOA1*, and *LCAT*) in individuals representing the upper and lower 5th percentile of HDL-cholesterol levels in a population-based sample. The benefit of selecting individuals with extreme values of quantitative traits is to increase the chance of detecting rare variants that may not be present in individuals with commonly observed quantitative values. Cohen et al. (2004) concluded that multiple rare nonsynonymous sequence variants of the genes that are involved in HDL metabolism may be influencing the HDL-cholesterol levels (overrepresented in the low-HDL group). Their results supported mainly the "rare variant-common disease" hypothesis for the genes that they screened.

We used a similar approach as Cohen et al. (2004), but we chose to sequence the entire gene and flanking regions so that we could also identify the regulatory variants. After sequencing 95 Black individuals and 95 NHW individuals in the extremes of HDL-cholesterol levels, we

found 26 variants (25 single nucleotide changes and 1 microsatellite; the latter was difficult to precisely determine using the sequencing strategy). This is more than the 15 variants (14 single nucleotide changes and 1 microsatellite) identified by Fullerton et al. (2002) and reported in the Seattle database (24 individuals from each of 3 populations; American Blacks, European NHWs, and American NHWs). Comparison of the variants identified in both studies is depicted in Table 21 including allele frequencies.

Table 21. *APOA2* variants and their allele frequencies: Comparison between our study and Seattle SNPs Database/Fullerton et al. (2002). The microsatellite marker is not included in this table, but was observed in all populations.

Site	Minor allele	Minor allele frequency				
		JD-pop*	ND-pop*	RD-pop*	NHWs**	Blacks**
155	C	0.042	0.250	0.326	0.321	0.063
159	G	-	-	-	0.011	-
201	A	-	-	0.021	-	-
336	G	-	-	-	0.005	-
470	A	-	-	-	-	0.005
589	T	-	-	-	-	0.021
772	T	-	-	-	0.016	-
872	C	0.333	0.312	0.375	0.405	0.195
898	G	-	-	-	0.005	-
1055	T	-	-	-	-	0.042
1218	C	-	0.021	-	-	-
1569	C	-	-	-	-	0.005
1638	T	-	-	-	-	0.027
2038	C	-	0.083	0.167	0.168	0.095
2085	T	-	0.021	-	0.005	-
2115	A	-	0.021	0.021	0.016	-
2218	C	-	-	-	0.005	-
2233	T	0.021	0.146	0.167	0.084	0.005
2298	T	-	-	-	-	0.026
2547	T	-	-	-	-	0.005
2818	T	-	0.083	0.167	0.168	0.095
2868	A	0.021	0.146	0.167	0.084	0.005
2994	T	-	0.021	0.146	0.163	0.053
3027	C	-	0.104	0.283	0.200	0.095
3092	G	-	0.104	0.292	0.197	0.083
3208	A	-	0.062	0.125	0.038	-
3251	G	-	-	-	0.048	-

*SeattleSNP Database - JD: Jackson,MS , ND: North_Karelia,Finland, RD: Rochester,MN
** From our study – NHWs: San Louis Valley, CO, Blacks: Benin City, Nigeria

Comparisons of sequencing data and conclusions between our study and their study are summarized as follows:

1. There were 2 variants reported in the Seattle database that were not observed in our sequencing sample. Those variants were a G>A variant at position 201 (MAF=0.02, only in American NHWs), and a G>C variant at position 1218 (MAF=0.002, only in European NHWs). The MAF of these variants suggests that only a single copy of the minor allele was observed in the entire white (European and American) population screened. Given the fact that our NHW and Black sample size was ~2-4 times larger than their sample size, we would expect to observe these variants in our sample. Therefore, we think that they may be sequencing artifacts. Alternatively, those variants may be genuine and the reason we did not observe them in our population sample could be due to sampling difference (subjects chosen in our study on the basis of extreme HDL levels versus subjects chosen in their study regardless of health status).
2. We detected a total of 12 previously unreported variants in our sequencing sample; 6 were present in NHWs and 6 were present in Blacks. The observation of these new variants could be due to a larger sample size allowing us to have a higher detection rate, or could be due to their association with extreme HDL levels. Of 7 variants in which only a single minor allele in the sequencing sample was observed (4 in NHWs and 3 in Blacks), one (2085C>T) was confirmed using TaqMan genotyping of the entire population and one (1569T>C) was confirmed using RFLP on selected samples. We failed to confirm one variant (336T>G) using RFLP, and considered it to be a sequencing artifact. Four variants remain to be confirmed (470G>A, 898A>G, 2218A>C, 2547G>T).

3. Previously, Fullerton et al. (2002) observed all population-specific variants as single-allele copies in their sequencing sample for the APOA2 gene. However, this was not the case in our study. All the variants that we observed as only a single allele in NHWs, however, were specific to that population. Fullerton et al. (2002) did not detect any population-specific variants in American Blacks. We were able to detect a total of 7 population-specific variants (MAF ranging from 0.005 to 0.042).
4. Fullerton et al. (2002) found that variant 872T>C had the highest frequency (MAF=0.333) among American Blacks, followed by 155G>C (MAF=0.042). We also found 872T>C as the most common variant in African Blacks (MAF=0.195). This SNP is functional, affecting known transcription binding site in the proximal promoter element D. The second most common variants in African Blacks were 2038G>C, 2818C>T, and 3027T>C (MAF=0.095). Fullerton et al. found 872T>C to be out of HWE in American Blacks. The SNP did not show deviation from HWE in African Blacks in our study. Given the fact that the MAF of this SNP in American Blacks falls in between what is observed in NHWs and African Blacks, deviation from HWE in American Blacks may be due to the migration and admixture.
5. Fullerton et al. (2002) observed a deficit of *APOA2* genetic variation in American Blacks as compared with American NHWs and European NHWs. They identified a total of 5 variants in American Blacks versus 15 variants in NHWs both including the microsatellite repeat marker. We identified a total of 17 variants in African Blacks versus 19 variants in NHWs. Even though the number of variants was not low in African Blacks, we observed lower rates of heterozygosity. For each single nucleotide variant that was common to both NHWs and African Blacks in our study (n=9), African Blacks had

significantly lower heterozygosity for rare alleles and lower homozygosity for common alleles. All the variants that were present in both populations showed significant differences in genotype and allele frequency when compared between races ($p<0.0001$).

6. Fullerton et al. (2002) did not observe any variants in the coding region of the gene. However, we were able to detect one synonymous sequence variant (2298C>T, MAF=0.026) only observed in African Blacks. Because we see no amino acid-changing variants, we expect any HDL-associated variants to be regulatory (modulating gene expression or affecting RNA processing). Nine of the 15 variants identified by Fullerton et al. (2002) did fall in the intronic regions and the remaining 9 (60%) fell in flanking regions (3 in 5'-flanking region and 6 in 3'-flanking region). Similarly, we identified 16 of 25 (64%) variants (excluding the unconfirmed variant at position 336) in the flanking regions (8 in 5'-flanking region and 8 in 3'-flanking region).
7. Fullerton el al. (2002) found that both variants with highest frequency (variants at position 155G>C and 872T>C) fell in the 5'-flanking region (putative promoter region). We found several variants more frequent than 155G>C, ranging from 5% to 10% frequency in the 3'-flanking region, which is different than their observation.

Cohen et al. (2004) found an increase in rare, nonsynonymous variants for *ABCA1*, *APOA1*, and *LCAT* in the low HDL group versus the high HDL group. Of the 128 individuals (64 Whites, 64 Blacks) with low plasma levels of HDL-cholesterol, 21 (16%) had unique sequence variants (not present in the high HDL-cholesterol group) versus 3 (2%) of individuals with unique sequence variants in the high HDL group (not present in the low HDL-cholesterol group). They found that the number of synonymous variants was similar in high and low HDL

groups. There are no known nonsynonymous changes in *APOA2*, as confirmed in our study as well.

In NHWs, we found a significant difference of allele frequencies for 2233C>T/rs6413453 and 2868C>A/rs12721036 SNPs between the high- and low- HDL groups (p=0.038 for each). These two SNPs were found to be closely linked to each other in LD analysis. We observed 2 rare variants (excluding unconfirmed 336T>G) only in the low HDL group (898A>G and 2085C>T), and 1 rare variant only in the high HDL group (2218A>C). For 6 of 14 SNPs found in both groups, homozygosity was observed only in the low HDL group (3251A>G, 2038G>C, 2233C>T, 2818C>T, 2868C>A, and 2994C>T), although heterozygotes were seen in both high and low HDL groups. There was no variant that was observed in homozygosity in only the high HDL group. Of the 48 individuals with low plasma levels of HDL, 6 (12.5%) had either rare variants (excluding the unconfirmed variant at position 336) not present in the high HDL group (N=2) or carried between 1 and 4 variants that were observed in the homozygous state only in the low HDL group (N=4). In contrast, of the 47 individuals with high plasma levels of HDL, only 1 subject (2.1%) had a rare variant not present in the low HDL group, and none carried variants that were homozygous only in the high HDL group. In Blacks, we found no significant difference of allele frequencies of *APOA2* variants between the high- and low- HDL groups. Three rare variants were observed only in the low HDL group (2233C>T, 2547G>T, 2868C>A) and 2 rare variants were observed only in the high HDL group (470G>A and 1569T>C). Eleven SNPs were present in both high HDL and low HDL groups, but homozygosity was observed only in the low HDL group for 6 variants (155G>C, 2038G>C, 2818C>T, 2818C>T, 2994C>T, 3027T>C, 3092A>G) and observed only in the high HDL group for 1 variant (872T>C). Of the 47 individuals with low plasma levels of HDL, 4 (8.5%) had either rare variants not present in

the high HDL group (N=3) or carried 5 variants that were observed in the homozygous state only in the low HDL group (N=1). Of the 48 individuals with high plasma levels of HDL, 3 (6.3%) had either a rare variant not present in the low HDL group (N=2), or carried a variant that was homozygous only in the high HDL group (N=1).

This observation of increased number of minor alleles of *APOA2* variants (increased heterozygosity for rare variants or increased homozygosity for common variants) needs confirmation in larger population-based samples. This observation suggests a cumulative effect of several regulatory *APOA2* variants on plasma HDL concentration. A similar observation was made by Cohen et al. (2004) for an accumulation of damaging nonsynonymous alleles of three other genes.

We performed a preliminary analysis using only 9 variants that were screened in NHWs (n=624, 8 variants) and/or Blacks (n=788, 5 variants) with TaqMan SNP genotyping assays. No association with HDL levels was found in the Black males or females. For the variants that were screened in NHWs, we found significant association in only females for variants 2233C>T/rs6413453 (adjusted p=0.028 for square root transformation of HDL levels) and 3251A>G (adjusted p=0.023 for square root transformation of HDL levels). However, we were not able to confirm the same association in males. Because there were only 3 females who were homozygous for the minor alleles of these 2 variants, we repeated our analysis in females by comparing wild-type allele homozygotes vs. minor allele heterozygotes+homozygotes. We still observed a slight significance for the association of rs6413453 (p=0.043), however there was no longer a significant association with 3251A>G (p=0.803). The 2233C>T variant is located at the 3' end of intron 3, close to the splice-site junction. The putative promoter variant

(872T>C/rs5082/-265T>C) that was found to be associated with other quantitative traits in prior studies did not show association with HDL levels in our study.

Due to time constraints and technical issues (TaqMan assays were not available for all variants, therefore more time-consuming methods such as RFLP and pyrosequencing will be performed), we were not able to complete genotyping of the remaining variants. When all variants are genotyped in the entire population, we will be able to determine the extent to which *APOA2* variants influence HDL levels in NHWs and Blacks.

BIBLIOGRAPHY

Alaupovic P, Knight-Gibson V, Wang V-S, Downs D, Koen E,. Brewer HB Jr., and Gregg RE. Isolation and characterization of an apoA-II-containing lipoprotein (LP-A-II-B complex) from plasma very low density lipoproteins of patients with Tangeir disease and type V hyperlipoproteinemia. *J Lipid Res* 1991;32: 9-19.

American Heart Association Website: AHA Recommendation for Cholesterol Levels. American Heart Association, Inc. Website accessed March 20, 2008. Copyright 2006. http://www.americanheart.org/presenter.jhtml?identifier=4500.

Assman G, Cullen P and Schulte H, The Münster Heart Study (PROCAM), results of follow-up at 8 years, *Eur Heart J* 1998;19:(suppl A) A2–A11.

Barbaras R, Puchois P, Pruchart JC, Pradines-Figueres A, Ailhaud G. Purification of an apolipoprotein A binding protein from mouse adipose cells. *Biochem J* 1990;269:767-73.

Berg K. Genetics of coronary heart disease and its risk factors. *Ciba Found Symp* 1987;130:14-33.

Blanco-Vaca F, Escola-Gil JC, Martin-Campos JM, and Julve J. Role of apoA-II in lipid metabolism and atherosclerosis: advances in the study of an enigmatic protein. *J Lipid Res* 2001;42: 1727-1739.

Blanco-Vaca F, Via DP, Yang C-Y, Massey JB, and Pownall HJ. Characterization of disulfide-linked heterodimers containing apolipoprotein D in human plasma lipoproteins. *J Lipid Res* 1992;33: 1785-96.

Borghini I., James RW, Blatter M-C, and Pomette D. Distribution of apolipoprotein E betweeen the free and apoA-II complexed forms in ver-low- and high-density lipoproteins: functional implications. *Biochim Biophys Acta* 1991;1083: 139-146.

Breslow LH. Apolipoprotein genetic variation and human disease. *Physiol Rev* 1988;68:85-132.

Brewer HB, Jr. Ronan R, Meng M, Bishop C. Isolation and characterization of apolipoproteins A-I, A-II, and J-IV. *Methods Enzymol* 1986;128:223-46.

Brewer HB, Lux SE, Ronan R, and John KM. Amino acid sequence of human apoLp-GlnII (apoA-II), an apolipoprotein isolated from the high density lipoprotein. *Proc Natl Acad Sci USA* 1972; 69: 1304-1308.

Brown GB, Zhao XQ and Cheung MC, Should both HDL-C and LDL-C be targets for lipid therapy? A review of current evidence, *J Clin Lipidol* 2007;1:88–94.

Bu X, Warden CH, Xia Y-R, De Meester C, Puppione DL, Teruya S, Lokensgard B, Daneshmand S, Brown J, Gray RJ, Rotter JI, and Lusis AJ. Linkage analysis of the genetic diterminants of high density lipoprotein concentrations and composition: evidence for involvement of the apolipoprotein A-II and cholesteryl ester transfer protein loci. *Hum Genet* 1994;93: 639-648.

Bunker CH, Ukoli FA, Matthews KA, Kriska AM, Huston SL, Kuller LH. Weight threshold and blood pressure in a lean black population. *Hypertension* 1995;26:616-623.

Bunker GH, Ukoli FA, Okoro FI, Olumu AB, Kriska AM, Huston SL, Markovic N, Kuller LH. Correlates of serum lipids in a lean black population. *Atherosclerosis* 1996;123;215-225.

Cambien F. Coronary Heart Disease and Polymorphisms in Genes Affecting Lipid Metabolism and Inflammation. *Curr Athero Rep* 2005;7:188-195.

Cardot P, Chambaz J, Claradas C, and Zannis VI. 1991. Regulation of the human apoA-II gene by the synergistic action of factors binding to the proximal and distal regulatory elements. *J Biol Chem* 1991; 266: 24460-24470.

Cardot P, Chambaz J, Kardassis J, Chambaz J, and Zannis VI. Factors participating in the liver-specific expression of the human apolipoprotein A-II gene and their significance for transcription. *Biochemistry* 1993;32: 9080-9093.

Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-sucleotide polymorhpisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004;74: 106-120.

Castelli WP, Garrison RJ, Wilson PWF, Abbott, Kalousdian S and Kannel WB, Incidence of coronary heart disease and lipoprotein cholesterol levels: the Framingham Study, *JAMA* 1986;256: 2835–2838.

Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 2004;305: 869-872.

Crawford DC, Yi Q, Smith JD, Shephard C, Wong M, Witrak L, Livingston RJ, Dieder MJ, and Nickerson DA. Allelic spectrum of the natural variation in *CRP*. *Hum Genet* 2006;119: 496–504.

Davidson MH and Toth PP. High-density lipoprotein metabolism: Potential therapeutic targets. *Am J Cardiol* 2007;100[suppl]:32N-40N.

Deeb SS, Takata K, Peng RL, Kajiyama G, and Albers JJ. A splice-junction mutation responsible for familial apolipoprotein A-II deficiency. *Am J Hum Genet* 1990;46: 822-827.

Doolittle MH, LeBoeuf R, Warden CH, L. Bee LM, and Lusis JJ. A polymorphism affecting apolipoprotein A-II translational efficiency determines high density lipoprotein size and composition. *J Biol Chem* 1990;265: 16380-16388.

Dupuy-Gorce, AM, Desmarais E, Vigneron S, Buresi C, Nicaud V, Evans A, Luc G, Arveiler D, Marques-Vidal P, Cambien F, Tiret L, Crastes de Pauler A, and Roizes G. DNA polymorphisms in linkage disequilibrium at the 3' end of the human APO AII gene: relationships with lipids, apolipoporteins and coronary heart disease. *Clin Genet* 1996;50: 191-198.

Eggerman TL, Hoeg JM, Meng MS, Tombragel A, Bojanovski D, Brewer HB, Jr. Differential tissue-specific expression of human apoA-I ans apoA-II. *J Lipid Res 1991*;32:821-828.

Fields LE, Burt VL, Cutler JA, Hughes J, Roccella EJ, Sorlie P. The burden of adult hypertension in the United States, 1999–2000: a rising tide. Hypertension. 2004;44:398–404.

Fullerton SM, Clark AG, Weiss KM, Taylor SL, Stengard JH, Salomaa V, Boerwinkle E, Nickerson DA. Sequence polymorphism at the human apolipoprotein AII gene (APOA2): unexpected deficit of variation in an African-American sample. *Hum Genet* 2002;111: 75-87.

Goldbourt U, Yaari JS and Medalie JH, Isolated low HDL cholesterol as a risk factor for coronary heart disease mortality: a 21-year follow-up of 8,000 men, *Arterioscler Thromb Vasc Biol* 1997;17:107–113.

Gordon DJ, Probstfield JL, Garrison RJ, Neaton JD, Castelli WP, Knoke JD, Jacobs DR Jr, Bangdiwala S, and Tyroler H. High-density lipoprotein cholesterol and cardiovascular disease, *Circulation* 1989;79:8–15.

Hamman RF, Marshall JA, Baxter J, Kahn LR, Mayer EJ, Orleans M, Murphy JR, Lezotte DC. Methods and prevalence of non-insulin dependent diabetes mellitus in a biethnic Colorado population. The San Luis Valley Diabetes Study. *Am Epidemiol* 1989;129:295-311.

Harris MR, Bunker CH, Hamman RF, Sanghera DK, Aston CE, Kamboh MI. Racial differences in the distribution of a low density lipoprotein receptor-related protein (LRP) polymorphism and its association with serum lipoprotein, lipid and apolipoprotein levels *Atherosclerosis* 1998; 137:187-195.

Hegele RA, Wang J, Harris SB, Brunt JH, Young TK, Hanley AJ, Zinman B, Connelly PW, Anderson CM. Variable association between genetic variation in the CYP7 gene promoter and plasma lipoproteins in three Canadian populations. *Atherosclerosis* 1997;154: 579-587.

Humphries SE. DNA polymorphisms of the apolipoprotein genes: Their use in the investigation of genetic component of hyperlipidaemia and atherosclerosis. *Atherosclerosis* 1988;72: 89-108.

Hussain MM and Zannis VI. Intracellular modification of human apolipoprotein (apoA-II) and sites of apoA-II synthesis: comparison of apoA-II with apoC-II and apoC-III isoproteins. *Biochemistry* 1990;29:209-217.

Ikewaki K, Zech LA, Kindt M, Brewer HB Jr, and Rader DJ. Apolipoprotein A-II production rate in a major factor regulating the distribution of apolipoprotein A-I among HDL subclasses ApA-I and LpA-I:A-II in normolipemic humans. *Arterioscler Thromb Vasc Biol* 1995;15: 306-312.

Jones DW, Chambless LE, Folsom AR, Heiss G, Hutchinson RG, Sharrett AR, Szklo M, Taylor HA Jr. Risk factors for coronary heart disease in African Americans: the Atherosclerotic Risk in Communities Study, 1987–1997. *Arch Intern Med* 2002;162:2565–2571.

Kalopissis A-D, Pastier D and Chambaz J. Apolipoprotein A-II: beyond genetic associations with lipid disorders and insulin resistance. *Curr Opin Lipidol* 2003;14;165-172.

Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS, Wahlstrand B, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* 2008; 40;2:189-197.

Knoblauch H, Bauerfeind A, Toliat MR, Becker C, Luganskaja T, Gunther UP, Rohe K, Schuster H, Junghans C, Luft FC, Nurnberg P, Reich JG (2004). Haplotypes and SNPs in 13 lopid-relevant genes explain most of the genetic variance in high-density lipoprotein and low-density lipoprotein cholesterol. *Hum Mol Genet* 2004;13:993-1004.

Ladias JA, Hadzopoulou-Claradas M, Kardassis D, Cardot P, Cheng J, Zannis VI, and Claradas C. Transcriptional regulation of human apolipoprotein genes apoB, apoCIII and apoAII by members of the steroid hormone receptor superfamily HNF4, ARP-1, EAR-2 and EAR-3. *J Biol Chem* 1992;267: 15849-15860.

Lagrost L, Dengremont C, Athias A, de Geitere C, Fruchart JC, Lallemant C, et al. Modulation of cholesterol efflux from Fu5AH hepatoma cells by the apolipoprotein contrent of high density lipoprotein particles. Particles containgin various proportions of apolipoproetians A-I and A-II. *J Biol Chem* 1995;270:12004-12009.

Lewis G and Rader D, New insights into the regulation of HDL metabolism and reverse cholesterol transport, *Circ Res* 2005;96: 1221–1232.

Li W-H, Tanimura M, Luo C-C, Datta S, and Chanc L. The apolipoprotein multigene family: biosynthesis, structure-function relationship and evolution. *J Lipid Res* 1988;29:245-271.

Lusis AJ, Mar R, Pajukanta P. Genetics of atherosclerosis. *Annu Rev Genomics Hum Genet* 2004;5:189-218.

Mahlberg FH, Rothblat GH. Cellular cholesterol efflux. Role of cell membrane kinetic pools and interaction with apolipoproteins AI, AII, and Cs. *J Biol Chem* 1992;267:4541-50.

Martin-Campos JM, Escola-Gil JC, Ribas V, and Blanco-Vaca F. Apolipoprotein A-II, genetic variation on chromosome 1q21-q24, and disease susceptibility. *Curr Opin Lipidol* 2004;15:247-253.

Maxfield FR and Wustner D, Intracellular cholesterol transport, *J Clin Invest* 2002;110:891–898.

Morabia A, Cayanis E, Costanza MA, Ross BM, Flaherty MS, Alvin GB, Das K, Gilliam TC. Association of extreme blood lipid profile phenotypic variation with 11 reverse cholesterol transport genes and 10 non-genetic cardiovascular disease risk factors. *Hum Mol Genet* 2003;12:2733-2743.

National Center for Health Statistics. *Health, United States, 2005.* With Chartbook on Trends in the Health of Americans. Hyattsville, Md: National Center for Health Statistics; 2005. Available at: http:// www.cdc.gov/nchs/data/hus/hus05.pdf. Accessed October 25, 2006.

NHLBI tabulation of mortality for 2003 for ischemic heart disease in Table GMWK210F in the NCHS Data Warehouse, July 2006. http://www.cdc.gov/nchs/datawh.htm. Accessed October 25, 2006.

Nicholls SJ, Tuzcu EM, Sipahi I, Grasso AW, Schoenhagen P, Hu T, Wolski K, Crowe T, Desai MY, Hazen SL, Kapadia SR and Nissen SE, Statins, high-density lipoprotein cholesterol, and regression of coronary atherosclerosis, *JAMA* 2007; 297: 499–508.

O'Meara JG, Kardia SLR, Armon JJ, Brown A, Boerwinkle E, Turner ST. Ethnic and Sex Differences in the Prevalence, treatment, and Control of Dyslipidemia Among Hypertensive Adults in the GENOA Study. *Arch Intern Med* 2004;164:1313-1318.

Rewers M, Shetterly SM, Hoag S, Baxter J, Marshall J, Hamman RF. Is the risk of coronary heart disease lower in Hispanics than in non-Hispanic whites? The San Luis Valley Diabetes Study. *Ethnicity Dis* 1993;3:44-54.

Ribeiro A, Pastier D, Kardassis D, Chambaz J, and Cardot P. Cooperative binding of upstream stimulatory factor and hepatic nuclear factor 4 drives the transcription of the human apolipoprotein A-II gene. *J Biol Chem* 1999;274: 1216-1225.

Robinson JG and Davidson MH, Investigational drugs targeting HDL-cholesterol metabolism and reverse cholesterol transport, *Future Lipidol* 2007;3: 285–301.

Rosamond W, et al. Heart Disease and Stroke Statistics-2007 Update: A Report From the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation* 2007;115:e69-e171.

Scott J, Knott TJ, Priestley LM, Robertson ME, Mann DV, Kostner G, Miller GJ, and Miller NE. High-density lipoprotein composition is altered by a common DNA polymorphism adjacent to apoprotein AII gene in man. *Lancet* 1985;1:771-773.

Segrest JP, Jones MK, de Loof H, Brouilette CG, Venkatachalapathi YV, and Anantharamiah GN. The amphipathic helix in the exchangable apolipoproteins: a review of secondary structure and function. *J Lipid Res* 1992;33: 141-166.

Seo DC, Torabi MR. Racial/ethnic differences in body mass index, morbidity and attitude toward obesity among U.S. adults. *J Natl Med Assoc* 2006;98:1300 –1308.

Shadrina MI, Slominskii PA, Perova NV, Limborskaia SA. The apolipoprotein AII gene is not a risk factor for development ischemic heart disease in the Moscow population. *Genetika* 1997;33:1316-1318.

Smith SC, Allen J, Blair SN. AHA/ACC guidelines for secondary prevention for patients with coronary and other atherosclerotic vascular disease: 2006 update. *Circulation* 2006; 113: 2363-2372.

Stampfer MJ, Sacks FM, Salvin S, Willett WC and Hennekens CH, A prospective study of cholesterol, apolipoproteins, and the risk of myocardial infarction, *N Engl J Med* 1991;325: 373–381.

Tailleux A, Duriez P, Fruchart J-C, Clavey V. Apolipoprotein A-II, HDL metabolism and atherosclerosis. *Atherosclerosis* 2002;164: 1-13.

Theret N, Delbart C, Aguie G, Fruchart JC, Vassaux G, Ailhaud G. Cholesterol efflux from adipose cells is coupled to diacylglycerol production and protein kinase C activation. *Biochem Biophys Res Commun* 1990;173:1361-1368.

Toth PP. Reverse cholesterol transport: high-density lipoprotein's magnificent mile, *Curr Atheroscler Rep* 2003;5:386–393.

Toth PP. Reducing cardiovascular risk by targeting high-density lipoprotein cholesterol, *Curr Atheroscler Rep* 2007;9:81–88.

Toth PP and Gotto AM, High-density lipoprotein cholesterol. In: Gotto AM and Toth PP, Editors, *Comprehensive Management of High Risk Cardiovascular Patients* 2006;295–339.

Vaziri, ND. Molecular Mechanisms of Lipid Disorders in Nephrotic Syndrome. *Kidney International* 2003;63:1964–1976.

Vu-Dac N, Schoonjans K, Kosykh V, Dallongeville J, Heyman RA, Staels B, and Auwerx J. Retinoids increase human apolipoprotein A-II expression through activation of the retinoid X receptor but not the retinoic acid receptor. *Mol Cell Biol* 1996;16: 3350-2260.

Wang Y, Barbacioru C, Hyland F et al. Large scale real-time PCR validation on gene expression measurements from two commercial long-oligonucleotide microarrays. *BMC Genomics* 2006;7:59.

Warnick GR, Benderson JM, Albers JJ. Quantitation of high density lipoprotein subclasses after separation by dextran sulfate and magnesium precipitation. *Clin Chem* 1982;28:1574.

Weber JL, and May PE. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 1989;44:388-396.

Weisgraber KH, and Mahley RW. Apoprotein (E-A-II) complex of human plasma lipoproteins. Characterization of this mixed disulfide and its identification in a high density lipoprotein subfraction. *J Biol Chem* 1978;253: 6281-6288.

Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, Heath SC, Timpson NJ, Najjar SS, Stringham HM, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 2008; 40;2:161-169.

Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, McQueen M, Budaj A, Pais P, Varigos J, Lisheng L; INTERHEART Study Investigators. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): Case-control study. *Lancet* 2004;364:937–952.

Zannis VI, Kan H-Y, Kritis A, Zanni E, and Kardassis D. Transcrptional regulation of the human apolipoprotein genes. *Front Biosci* 2001;6:456-504.